

THESIS / THÈSE

MASTER EN SCIENCES MATHÉMATIQUES

Généralisation à la programmation semi-définie de la méthode de suivi de chemin établie en programmation linéaire.

Charles, Catherine

Award date:
1998

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

FACULTÉS UNIVERSITAIRES N.D. DE LA PAIX
NAMUR
FACULTÉS DES SCIENCES

Généralisation
à la programmation semi-définie
de la méthode de suivi de chemin
établie en programmation linéaire

Mémoire présenté pour l'obtention du grade
de Licencié en Sciences
mathématiques
par

Promoteur : J.J.Strodiot

Catherine CHARLES

Année académique 1997-1998

Au terme de ce mémoire, je tiens à remercier mon promoteur, monsieur Strodiot, qui m'a suivie tout au long de ce travail de fin d'études et qui a toujours été disponible à mon égard. J'y associe tous les professeurs qui ont contribué généreusement à ma formation. Je souhaite également témoigner ma gratitude à mon entourage pour ces quatre belles années passées aux facultés.

Résumé

La programmation semi-définie consiste à minimiser une fonction linéaire en tenant compte d'inégalités matricielles. Elle recouvre de nombreux problèmes standards, dont la programmation linéaire. De plus, la plupart des résultats ainsi qu'un grand nombre de procédés de résolution utilisés en programmation linéaire s'y étende. Plus particulièrement, ce mémoire étudie la généralisation à la programmation semi-définie de la méthode primale-duale de points intérieurs et de suivi de chemin établie en programmation linéaire. Celle-ci engendre des points suivant de manière approximative une trajectoire spécifique. Ce chemin, situé à l'intérieur du domaine admissible, mène à une solution optimale du problème primal-dual initial. L'extension à la programmation semi-définie de différents algorithmes basés sur cette méthode et de leurs propriétés de convergence s'effectue grâce à un artifice de calcul se basant sur une transformation du problème primal-dual initial. La généralisation de la dualité et de procédés déterminant l'admissibilité du problème (avec obtention d'un itéré de départ) se réalise par l'intermédiaire de la théorie de la dualité en programmation conique convexe. Similairement à la programmation linéaire, cette méthode possède une complexité polynomiale et s'avère performante d'après les utilisateurs.

Abstract

Semidefinite programming consists in minimizing a linear function subject to matrix inequalities. It unifies several standard problems like linear programming. Moreover, the most of the results and a great number of resolution processes used in linear programming apply in semidefinite programming. Particularly, this paper studies the generalization to semidefinite programming of the primal-dual interior-point path-following method, set up in linear programming. This method generates points which follow roughly a specific path. This path, situated inside the feasible set, leads to an optimal solution of the initial primal-dual problem. The extension to semidefinite programming of different algorithms based on this method and of their convergence properties is done thanks to a transformation of the initial primal-dual problem. The generalization of the duality and of processes determining the admissibility of the problem or an initial interior point is realised by results from the duality in closed conic convex programming. As in linear programming, this method has a polynomial worst-case complexity, and performs very well in practice according to the users.

Introduction

Historiquement, la programmation linéaire, qui consiste à minimiser une fonction linéaire en tenant compte de contraintes linéaires, a été à l'origine du développement d'algorithmes en optimisation. L'algorithme du simplexe introduit dans les années 1940 ainsi que les différentes méthodes de résolution de ce problème développées par la suite ont aidé de nombreux professionnels de l'économie et de la finance.

Depuis peu, de nombreux mathématiciens s'intéressent à un problème plus général, la programmation semi-définie. Celle-ci recouvre une grande classe de problèmes standards, dont la programmation linéaire, quadratique et convexe et trouve ainsi énormément d'applications, entre autres en optimisation combinatoire non convexe et en théorie du contrôle. Puisque la programmation semi-définie est considérée comme une extension de la programmation linéaire, elle reste dans la continuité de ce qui a déjà été réalisé. Ainsi, beaucoup de résultats et de méthodes de points intérieurs pour la programmation linéaire ont pu être généralisés à la programmation semi-définie. Tout comme la programmation linéaire, ces méthodes possèdent une complexité algorithmique polynomiale et s'avèrent très efficaces.

Ce mémoire, basé principalement sur la thèse de doctorat de J.F. Sturm (septembre 1997), propose d'étudier une méthode particulière de résolution de la programmation semi-définie, méthode généralisée de la programmation linéaire : la méthode primale-duale de points intérieurs et de suivi de chemin.

Le premier chapitre débute par la définition du problème de programmation semi-définie. Trois formulations sont présentées, mettant en évidence soit la généralisation de la programmation linéaire, soit le cas particulier que représente la programmation semi-définie par rapport à la programmation conique convexe. Tout au long de ce mémoire, nous tirerons profit autant que possible des avantages de chacune de celles-ci. Afin de mieux rendre compte de l'utilité de la programmation semi-définie et de la raison de son étude, plusieurs applications sont reprises dans ce chapitre.

Le deuxième chapitre généralise la théorie de la dualité de la programmation linéaire à la programmation conique convexe fermée, et de ce fait à la programmation semi-définie. Ce chapitre est essentiel puisque ce mémoire étudie une méthode primale-duale. L'association d'un problème de programmation conique convexe dual au problème de programmation conique convexe de départ permet de caractériser les notions d'admissibilité et de non-

admissibilité d'un problème. Cette caractérisation rendra possible dans le dernier chapitre la détermination du caractère admissible du problème de programmation conique convexe initial.

Tout comme en programmation linéaire, la dualité permet également de calculer des bornes et d'obtenir des informations sur la valeur optimale recherchée. Ainsi, quelle que soit l'admissibilité des problèmes primal et dual de départ, nous généralisons à la programmation conique convexe la relation de dualité faible entre les valeurs optimales primales et duales. Le théorème de dualité forte étendu à la programmation conique convexe nécessite, quant à lui, une hypothèse plus restrictive, celle de l'admissibilité forte primale.

Le troisième chapitre généralise, sous l'hypothèse d'admissibilité forte primale et duale, la méthode primale-duale de points intérieurs et de suivi de chemin de la programmation linéaire à la programmation semi-définie. Les raisons du choix d'une telle méthode sont explicitées dans ce chapitre. Cette extension s'effectue grâce à un artifice de calcul se basant sur une transformation du problème primal-dual initial. L'algorithme à petits pas, l'algorithme prédicteur-correcteur et l'algorithme à longs pas relèvent du même principe que leurs analogues en programmation linéaire et sont donc d'efficacité semblable. Nous terminons ce chapitre par une discussion sur les nombreuses directions de descente qu'une telle méthode peut envisager en programmation semi-définie.

Le quatrième chapitre précise la solution optimale vers laquelle convergent les algorithmes basés sur la méthode primale-duale de points intérieurs et de suivi de chemin développée dans le chapitre précédent. Cette convergence vers le centre analytique des ensembles optimaux primal et dual généralise celle de la programmation linéaire. Ensuite, nous détaillons un algorithme de type prédicteur-correcteur découlant de cette méthode. Nous démontrons la convergence globale de cet algorithme ainsi que la convergence superlinéaire des itérés.

La méthode primale-duale de points intérieurs et de suivi de chemin, développée dans les chapitres précédents pour la résolution d'un problème de programmation semi-définie, présuppose l'admissibilité forte primale et duale ainsi que la connaissance d'un point de départ strictement admissible. Dans la pratique, cependant, à la vue d'un problème primal-dual, nous ne sommes en général pas capable de juger de son admissibilité. De plus, il n'est pas toujours évident de trouver un point initial intérieur. Le dernier chapitre propose différentes méthodes pour pallier à ces deux inconvénients. En fait, d'une part, pour déterminer un point de départ strictement admissible, nous généralisons, de la programmation linéaire à la programmation conique convexe fermée, la méthode basée sur la self-dualité, et, à la programmation semi-définie, celle du grand M. D'autre part, pour connaître le caractère admissible des problèmes primal et dual initiaux, nous étendons également la méthode de Ye, Todd et Mizuno de la programmation linéaire à la programmation conique convexe fermée.

Chapitre 1

La programmation semi-définie

Nous commençons par définir la programmation semi-définie. Cependant, vu le caractère récent de ce problème, la formulation de celui-ci n'est pas encore uniformisée dans la littérature. C'est pourquoi, nous dégageons ici trois écritures distinctes. Celles-ci sont exploitées au mieux tout au long de ce mémoire.

Ensuite, différentes applications illustrent ce problème d'optimisation et valident son analyse.

1.1 Trois formulations du problème de programmation semi-définie

Un problème de programmation semi-définie consiste à minimiser une fonction linéaire en tenant compte d'inégalités matricielles. Un tel problème se note :

$$\inf c^T x \quad \text{s.c. } F(x) \succeq 0 \quad (1.1)$$

où $c, x \in R^m$, $F(x) = F_0 + \sum_{i=1}^m x_i F_i$, et $F_0, \dots, F_m \in C^{n \times n}$ et hermitiennes. La contrainte $F(x) \succeq 0$ signifie que $F(x)$ est semi-définie positive.

Cette formulation du problème apporte deux éléments de réponse au pourquoi de l'étude de ce sujet.

En effet, d'une part, la programmation semi-définie peut être considérée comme une extension de la programmation linéaire. Afin de mieux s'en rendre compte, considérons le problème de programmation linéaire suivant :

$$\inf c^T x \quad \text{s.c. } Ax + b \geq 0 \quad (1.2)$$

où $c, x \in R^m$, $b \in R^n$, $A = [a_1 \dots a_m] \in R^{n \times m}$.

En posant $F(x) = \text{diag}(Ax + b)$, c'est-à-dire $F_0 = \text{diag}(b)$, $F_i = \text{diag}(a_i)$ pour $i = 1 \dots m$, et en nous souvenant de la propriété suivante, $v \geq 0 \iff \text{diag}(v) \succeq 0$, le problème

de programmation linéaire (1.2) peut être considéré comme un problème de programmation semi-définie. Nous pouvons également voir la programmation semi-définie comme un problème de programmation linéaire semi-infini puisque l'inégalité matricielle $F(x) \succeq 0$ est équivalente à un ensemble infini de contraintes linéaires sur $x : \forall z \in R^n, z^T F(x) z \geq 0$. Dès lors, il ne faut pas s'étonner du parallélisme entre la théorie de la programmation semi-définie et celle de la programmation linéaire ainsi que de la généralisation de beaucoup d'algorithmes de la programmation linéaire à la programmation semi-définie. Les méthodes de résolution pour la programmation semi-définie, certaines développées dans des chapitres ultérieurs, sont légèrement plus compliquées que celles de la programmation linéaire mais tout aussi efficaces. Nous constaterons cependant certaines différences importantes entre ces deux problèmes, comme des résultats de dualité plus faibles,...

D'autre part, la contrainte de (1.1) est convexe. Ainsi, un problème de programmation semi-définie est un problème d'optimisation convexe. C'est pourquoi la programmation semi-définie recèle un nombre important d'applications.

Par conséquent, la programmation semi-définie est fortement prisée de par son efficacité ainsi que par la grande classe de problèmes qu'elle embrasse.

Comme souligné au début de ce chapitre, la théorie de la programmation semi-définie étant assez récente, la notation de ce problème n'est pas encore uniformisée dans la littérature. Si l'on considère le problème comme étant une généralisation de la programmation linéaire, le problème se note (1.1). C'est le cas de Lieven Vandenberghe et de Stephen Boyd dans [15]. Son avantage est la simplicité et son facile raccord à la programmation linéaire.

Si, par contre, l'on considère le problème comme étant un cas particulier de la programmation conique convexe, son écriture est différente. Envisageons cette écriture qui a été adoptée, entre autres par Sturm dans [1], principale référence de ce mémoire.

Tout d'abord, introduisons quelques définitions et notations nécessaires à cette formulation.

Soit $R^{\bar{n} \times \bar{n}}$, l'espace des matrices réelles carrées de dimension \bar{n} . Définissons sur cet espace le produit interne standard $X \bullet Y = \text{tr} X^T Y$ où tr désigne la trace. Il suit que $X \perp Y \iff X \bullet Y = 0$. Si nous généralisons ce produit à $C^{\bar{n} \times \bar{n}}$, nous obtenons $X \bullet Y = (\text{Re} X \bullet \text{Re} Y) + (Im X \bullet Im Y) = \text{Re } \text{tr} Y^H X$. Soit $\mathcal{H}_{\bar{n}}$, l'espace réel linéaire des matrices hermitiennes $\bar{n} \times \bar{n}$. Nous notons $X \succeq 0$ ($\succ 0$) si $X \in \mathcal{H}_{\bar{n}}$ et X (semi-)définie positive.

Ensuite, introduisons la formulation d'un problème de programmation semi-définie considérée comme cas particulier de la programmation conique convexe. Nous avons :

$$\inf C \bullet X \quad \text{s.c. } A_i \bullet X = b_i, i = 1 \dots m \text{ et } X \succeq 0 \quad (1.3)$$

où $C, X \in \mathcal{H}_{\bar{n}}$, $A_i \in \mathcal{H}_{\bar{n}} \forall i = 1 \dots m$, et $b \in R^m$.

Cette notation dévoile aussi de façon évidente la généralisation de la programmation linéaire à l'ensemble des matrices hermitiennes.

Montrons que le problème de programmation semi-définie exprimé en (1.3) est équivalent à celui exprimé en (1.1). A priori, ceci semble peu direct. Pour contourner cette difficulté, nous nous raccrochons au fait qu'en programmation semi-définie, nous traitons des méthodes primales-duales, fournissant une solution optimale primale et une solution optimale duale, et au caractère involutif de la dualité (voir théorème 2.1). Ainsi, il nous suffit d'exhiber la similarité entre le problème primal, (1.3), et le problème dual de (1.1).

Le dual de (1.1) vaut

$$\sup \operatorname{Re} \operatorname{tr}(-F_0)Z \quad \text{s.c.} \quad \operatorname{Re} \operatorname{tr} F_i Z = c_i \quad i = 1 \dots m, \quad Z \succeq 0$$

qui est équivalent au problème d'optimisation suivant

$$-\inf F_0 \bullet Z \quad \text{s.c.} \quad F_i \bullet Z = c_i \quad i = 1 \dots m, \quad Z \succeq 0.$$

En posant $C = F_0 \in \mathcal{H}_{\bar{n}}$; $A_i = F_i \in \mathcal{H}_{\bar{n}}$; $b_i = c_i \in R$ pour tout i et $X = Z \in \mathcal{H}_{\bar{n}}$, nous trouvons le problème suivant :

$$-\inf C \bullet X \quad \text{s.c.} \quad A_i \bullet X = b_i \quad i = 1 \dots m \text{ et } X \succeq 0$$

qui est bien (1.3).

Ensuite, nous nous attachons à justifier le fait que la programmation semi-définie soit un cas particulier de la programmation conique convexe. Dans ce but, définissons tout d'abord un cône convexe et un problème de programmation conique convexe. Un cône convexe non vide est un ensemble \mathcal{K} tel que $\{0\} \in \mathcal{K}$ et $\alpha(\mathcal{K} \oplus \mathcal{K}) = \mathcal{K}$, $\forall \alpha > 0$, où $T \oplus T' = \{z \in R^n | z = x + y, x \in T, y \in T'\}$. Un problème de programmation conique convexe est un problème d'optimisation dont la fonction objectif est linéaire et donc les contraintes représentent l'intersection d'un espace affine avec un cône convexe. Il est donc formulé de la manière suivante :

$$\inf C \bullet X \quad \text{s.c.} \quad X \in (B + \mathcal{A}) \cap \mathcal{K} \tag{1.4}$$

où \mathcal{A} est un sous-espace linéaire de $\mathcal{H}_{\bar{n}}$, \mathcal{K} est un cône convexe dans $\mathcal{H}_{\bar{n}}$ et $C, B \in \mathcal{H}_{\bar{n}}$. Souvent, pour fixer la théorie, on suppose $C \in \mathcal{A}$ et $B \in \mathcal{A}^\perp$. Nous effectuons cette hypothèse. Par souci de facilité, on note ce problème $CP(B, C, \mathcal{A}, \mathcal{K})$.

Il nous reste à retrouver dans le problème (1.3) un problème de programmation conique convexe de la forme (1.4). Soit $B \in \mathcal{H}_{\bar{n}}$ tel que $A_i \bullet B = b_i$ pour $i = 1 \dots m$. Si B n'existe pas, le système d'égalités $A_i \bullet X = b_i$ pour $i = 1 \dots m$ est inconsistent. Définissons $\mathcal{A} \subseteq \mathcal{H}_{\bar{n}}$ tel que $\mathcal{A} = \{Y \in \mathcal{H}_{\bar{n}} | A_i \bullet Y = 0 \quad i = 1 \dots m\}$. Le problème (1.3) peut se noter alors

$$\inf C \bullet X \quad \text{s.c.} \quad X \in (B + \mathcal{A}) \cap \mathcal{H}_{\bar{n}}^+ \tag{1.5}$$

où $\mathcal{H}_{\bar{n}}^+$ est l'ensemble des matrices $\bar{n} \times \bar{n}$ hermitiennes semi-définies positives. Vu que $\mathcal{H}_{\bar{n}}^+$ est un cône convexe, notre thèse est prouvée.

La troisième formulation du problème de programmation semi-définie reste proche de (1.3), cas particulier de la programmation conique convexe. La petite différence réside dans le fait que les variables sont vectorielles et non plus matricielles. Cette troisième écriture facilitera grandement certains raisonnements de ce mémoire, vu que nous travaillerons sur des vecteurs, outils que nous connaissons bien.

Nous définissons x comme vecteur des coordonnées de X dans une base de $\mathcal{H}_{\bar{n}}$. En effet, nous savons que $\mathcal{H}_{\bar{n}} = \mathcal{S}_{\bar{n}} + i\mathcal{S}_{\bar{n}}^\perp$ où $\mathcal{S}_{\bar{n}}$ est l'espace linéaire des matrices $\bar{n} \times \bar{n}$ réelles symétriques et que $m := \dim \mathcal{S}_{\bar{n}} = \bar{n} + (\bar{n} - 1) + \dots + 1 = \frac{\bar{n}(\bar{n}+1)}{2}$. Soit $(U_j)_{j=1\dots m}$ une base orthonormale fixée de $\mathcal{S}_{\bar{n}}$. De la même manière, $\dim \mathcal{S}_{\bar{n}}^\perp = (\bar{n} - 1) + (\bar{n} - 2) + \dots + 1 = \frac{\bar{n}(\bar{n}-1)}{2} = \bar{n}^2 - \frac{\bar{n}(\bar{n}+1)}{2}$ où $\bar{n}^2 = \dim \mathcal{H}_{\bar{n}}$. Soit $(U_j)_{j=m+1\dots \bar{n}^2}$ une base orthonormale fixée de $\mathcal{S}_{\bar{n}}^\perp$. Donc, $(U_j)_{j=1\dots \bar{n}^2}$ forme une base orthonormale de l'espace $\mathcal{H}_{\bar{n}}$. Etant donné $X \in \mathcal{H}_{\bar{n}}$, nous pouvons alors définir, comme souhaité, $x = \text{vec}_H(X)$, vecteur des coordonnées de X dans la base $(U_j)_{j=1\dots \bar{n}^2}$ appartenant à $R^n = R^{\bar{n}^2}$. Ainsi, nous avons

$$\begin{aligned} x_j &= U_j \bullet X & j = 1\dots m \\ &= (iU_j) \bullet X & j = m+1\dots n. \end{aligned}$$

Suite à cette définition, réexprimons notre problème de programmation semi-définie (1.5) comme cas particulier de la programmation conique convexe. Nous obtenons

$$\inf c^T x \quad \text{s.c. } x \in (b + \mathcal{A}) \cap \mathcal{H}_{\bar{n}}^+. \quad (1.6)$$

En effet, les problèmes (1.5) et (1.6) sont équivalents car

$$\begin{aligned} C \bullet X &= \left(\sum_{j=1}^n c_j U_j \right) \left(\sum_{k=1}^n x_k U_k \right) \\ &\quad \text{où } c = (c_j)_{j=1\dots n} = \text{vec}_H(C) \text{ et } x = (x_j)_{j=1\dots n} = \text{vec}_H(X) \\ &= \left(\sum_{j=1}^n c_j x_j \right) \text{ car } (U_j)_{j=1\dots n} \text{ est une base orthonormale} \\ &= c^T x \\ \text{et } b &= \text{vec}_H(B). \end{aligned}$$

Nous laissons l'écriture de \mathcal{A} et de $\mathcal{H}_{\bar{n}}^+$ comme telle par simplicité de notation. De nouveau, les hypothèses simplificatrices $c \in \mathcal{A}$ et $b \in \mathcal{A}^\perp$ restent valables ainsi que la notation concise du problème (1.6), $CP(b, c, \mathcal{A}, \mathcal{H}_{\bar{n}}^+)$.

Ces trois formulations différentes d'un problème de programmation semi-définie seront utilisées pertinemment tout au long de ce mémoire. Nous tirerons profit autant que possible des avantages de chacune.

1.2 Applications de la programmation semi-définie

Afin de mieux rendre compte de l'utilité de la programmation semi-définie, de donner une idée du problème traité, et afin de remarquer les liens qui existent entre des applications de domaines différents, certaines applications sont reprises ci-dessous. Nous utilisons dans cette section principalement la notation (1.1) d'un problème de programmation semi-définie.

1.2.1 Programmation quadratique avec contraintes quadratiques

Un problème de programmation quadratique avec contraintes quadratiques a la forme générale suivante :

$$(QCQP) : \quad \inf f_0(x) \quad \text{s.c.} \quad f_i(x) \leq 0 \quad i = 1 \dots L$$

où $f_i(x) = (A_i x + b)^T (A_i x + b) - c_i^T x - d_i$. Grâce au théorème de Schur, nous pouvons écrire une contrainte quadratique de la forme $(Ax + b)^T (Ax + b) - c^T x - d \leq 0$ où $x \in R^k$ comme ceci :

$$\begin{pmatrix} I & Ax + b \\ (Ax + b)^T & c^T x + d \end{pmatrix} \succeq 0.$$

Les éléments du membre de gauche de cette inégalité ne dépendent que de manière affine des éléments du vecteur x et donc le membre de gauche peut être exprimé comme suit :

$$F(x) = F_0 + x_1 F_1 + \dots + x_k F_k \succeq 0$$

où $F_0 = \begin{pmatrix} I & b \\ b^T & d \end{pmatrix}$, $F_i = \begin{pmatrix} 0 & a_i \\ a_i^T & c_i \end{pmatrix} \quad i = 1 \dots k$, et $A = [a_1 \dots a_k]$. Le problème de programmation quadratique avec contraintes quadratiques (QCQP) devient donc un problème de programmation semi-définie en (x, t) où t est une nouvelle variable introduite appartenant à R ,

$$\inf t \quad \text{s.c.} \quad \begin{pmatrix} I & A_0 x + b_0 \\ (A_0 x + b_0)^T & c_0^T x + d_0 + t \end{pmatrix} \succeq 0 \quad \text{et} \quad \begin{pmatrix} I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d_i \end{pmatrix} \succeq 0 \quad \text{pour } i = 1 \dots L.$$

1.2.2 Minimisation de la valeur propre maximale d'une matrice et de la norme matricielle

Supposons A une matrice symétrique.

Nous savons que $\lambda_{max} = \sup_{x \neq 0} \frac{x^T A x}{\|x\|_2^2}$. Par conséquent, λ_{max} est solution du problème

$$\min t \quad \text{s.c.} \quad \frac{x^T A x}{\|x\|_2^2} \leq t, \quad \forall x \neq 0,$$

qui est équivalent aux deux problèmes suivants :

$$\min t \quad \text{s.c.} \quad x^T A x - t \|x\|_2^2 \leq 0, \quad \forall x \neq 0,$$

$$\min t \quad \text{s.c.} \quad A - tI \preceq 0.$$

Calculer la valeur propre maximale de A , où A est symétrique, revient donc à résoudre le problème de programmation semi-définie suivant :

$$\min t \quad \text{s.c.} \quad tI - A \succeq 0$$

où $t \in \mathbb{R}$.

Nous savons également que $\|A\|_2 = \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$. Par conséquent, $\|A\|_2$ est solution du problème

$$\min t \quad \text{s.c.} \quad \frac{\|Ax\|_2^2}{\|x\|_2^2} \leq t^2, \quad \forall x \neq 0,$$

qui est équivalent aux deux problèmes suivants :

$$\min t \quad \text{s.c.} \quad x^T A^T A x - t^2 \|x\|_2^2 \leq 0, \quad \forall x \neq 0,$$

$$\min t \quad \text{s.c.} \quad A^T A - t^2 I \preceq 0.$$

Calculer la norme 2 d'une matrice symétrique A revient donc, grâce au théorème de Schur, à résoudre le problème de programmation semi-définie suivant

$$\min t \quad \text{s.c.} \quad \begin{pmatrix} tI & A \\ A^T & tI \end{pmatrix} \succeq 0.$$

1.2.3 Schéma de séparation par ellipsoïde

Pour séparer deux ensembles de points $\{x_1 \dots x_k\}$ et $\{y_1 \dots y_l\}$ dans \mathbb{R}^p , les classifications les plus simples utilisent des hyperplans. L'hyperplan défini par $a^T x + b = 0$ sépare ces deux ensembles si $a^T x_i + b \leq 0$ pour $i = 1 \dots k$ et $a^T y_j + b \geq 0$ pour $j = 1 \dots l$. Ces contraintes forment un ensemble d'inégalités linéaires en $a \in \mathbb{R}^p$ et $b \in \mathbb{R}$ résoluble par la programmation linéaire. Si ces deux ensembles ne peuvent être séparés par un hyperplan, nous pouvons essayer de les séparer par une surface quadratique. En d'autres mots, nous cherchons une fonction quadratique $f(x) = x^T A x + b^T x + c$ qui satisfasse

$$(x_i)^T A x_i + b^T x_i + c \leq 0 \quad \text{pour } i = 1 \dots k \quad (1.7)$$

$$(y_j)^T A y_j + b^T y_j + c \geq 0 \quad \text{pour } j = 1 \dots l. \quad (1.8)$$

Ces contraintes forment un ensemble d'inégalités linéaires en $A = A^T \in \mathbb{R}^{p \times p}$, $b \in \mathbb{R}^p$ et $c \in \mathbb{R}$. De nouveau, ce problème est résoluble grâce à la programmation linéaire.

Ne voulant pas en rester là, il nous est permis de pousser plus loin les restrictions sur la surface quadratique séparant les deux ensembles. Nous pouvons essayer par exemple de trouver un ellipsoïde qui contienne tous les x_i et aucun y_j . Pour satisfaire cela, en plus

des contraintes (1.7) et (1.8), nous devons imposer la condition $A \succ 0$. Le problème de programmation semi-définie associé est

$$\inf_0 \quad \text{s.c.} \quad \begin{pmatrix} -(x_i)^T A x_i - b^T x_i - c & 0 & 0 \\ 0 & (y_j)^T A y_j + b^T y_j + c & 0 \\ 0 & 0 & A - \epsilon I \end{pmatrix} \succeq 0$$

où ϵ est une constante strictement positive très petite. Si ce problème est admissible, il nous permet de trouver l'ellipsoïde demandé.

1.2.4 Optimisation combinatoire et non convexe

Comme nous allons le voir dans cet exemple-ci, la programmation semi-définie joue un rôle très utile dans l'optimisation combinatoire non convexe.

Soit le problème quadratique suivant :

$$\inf f_0(x) \quad \text{s.c.} \quad f_i(x) \leq 0, \quad i = 1 \dots l \quad (1.9)$$

où $f_i(x) = x^T A_i x + 2b_i^T x + c_i$, $i = 0 \dots l$. Aucune hypothèse n'est faite sur A_i , $\forall i = 0, \dots, l$. Les matrices A_i peuvent donc être indéfinies, ce qui complique fortement la résolution de ce problème. Plusieurs méthodes ont donc été développées pour répondre à cette question, entre autres l'algorithme "Branch and bound" qui requiert le calcul d'une borne inférieure assez précise de la valeur optimale de (1.9). Certains auteurs ont proposé de calculer une telle borne en résolvant le problème de programmation semi-définie ci-dessous :

$$\max t \quad \text{s.c.} \quad \begin{pmatrix} A_0 & b_0 \\ b_0^T & c_0 - t \end{pmatrix} + \tau_1 \begin{pmatrix} A_1 & b_1 \\ b_1^T & c_1 \end{pmatrix} \dots + \tau_l \begin{pmatrix} A_l & b_l \\ b_l^T & c_l \end{pmatrix} \succeq 0 \quad (1.10)$$

où $\tau_i \geq 0$ pour $i = 1 \dots l$.

Vérifions que la solution de ce problème de programmation semi-définie corresponde à une borne inférieure de la valeur optimale de (1.9).

Remarquons que si x satisfait les contraintes du problème non convexe (1.9), nous obtenons

$$f_i(x) = (x, 1) \begin{pmatrix} A_i & b_i \\ b_i^T & c_i \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} \leq 0 \quad i = 1 \dots l.$$

Soient t, τ_1, \dots, τ_l satisfaisant les contraintes de programmation semi-définie (1.10). Alors, nous avons pour x , point admissible de (1.9),

$$\begin{aligned} 0 &\leq (x, 1) \left(\begin{pmatrix} A_0 & b_0 \\ b_0^T & c_0 - t \end{pmatrix} + \tau_1 \begin{pmatrix} A_1 & b_1 \\ b_1^T & c_1 \end{pmatrix} + \dots + \tau_l \begin{pmatrix} A_l & b_l \\ b_l^T & c_l \end{pmatrix} \right) \begin{pmatrix} x \\ 1 \end{pmatrix} \\ &= f_0(x) - t + \tau_1 f_1(x) + \dots + \tau_l f_l(x) \\ &\leq f_0(x) - t \end{aligned}$$

car $\tau_i \geq 0$ et $f_i(x) \leq 0 \quad \forall i = 1 \dots l$. C'est pourquoi, pour tout x admissible de (1.9), nous avons $t \leq f_0(x)$ et rechercher le plus grand t est équivalent à rechercher le plus petit $f_0(x)$.

1.2.5 Théorie du contrôle

Considérons le problème de contrôle optimal suivant :

$$\frac{dx}{dt} = Ax(t) + Bu(t), \quad y(t) = Cx(t), \quad |u_i(t)| \leq |y_i(t)| \quad i = 1 \dots p \quad (1.11)$$

où $x(t) \in R^l$, $u(t)$ et $y(t) \in R^p$.

Une question importante dans la théorie du contrôle optimal est de trouver un ellipsoïde invariant, c'est-à-dire un ellipsoïde E vérifiant

$$\forall x, u \text{ satisfaisant (1.11), si } x(T) \in E \text{ alors } x(t) \in E \quad \forall t \geq T.$$

L'existence d'un tel ellipsoïde implique, par exemple, que toutes les solutions de (1.11) sont bornées. Par cette définition, nous pouvons dire qu'un ellipsoïde $E = \{x | x^T P x \leq 1\}$ où $P = P^T \succ 0$ est invariant

$$\begin{aligned} \iff V(t) = x(t)^T P x(t) \text{ est non croissante } \forall x, u \text{ satisfaisant (1.11)} \\ \iff \frac{dV(x(t))}{dt} \leq 0 \quad \forall x, u \text{ satisfaisant (1.11)} \\ \iff \frac{dV(x(t))}{dt} = (x(t), u(t)) \begin{pmatrix} A^T P + P A & P B \\ B^T P & 0 \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \end{pmatrix} \leq 0 \end{aligned} \quad (1.12)$$

pour tout x et u satisfaisant $y(t) = Cx(t)$ et $|u_i(t)| \leq |y_i(t)|$, $i = 1 \dots p$.

Exprimons les conditions $|u_i(t)| \leq |y_i(t)|$ $i = 1 \dots p$ comme des inégalités quadratiques $u_i(t)^2 - y_i(t)^2 \leq 0$, $i = 1 \dots p$:

$$\begin{aligned} u_i(t)^2 - y_i(t)^2 &= u_i(t)^2 - x(t)^T c_i^T c_i x(t) \\ &\quad \text{où } c_i \text{ est la } i\text{ème ligne de } C \text{ et } y(t) = Cx(t) \\ &= (x(t), u(t)) \begin{pmatrix} -c_i^T c_i & 0 \\ 0 & E_{ii} \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \end{pmatrix} \leq 0 \end{aligned} \quad (1.13)$$

pour $i = 1 \dots p$ et où E_{ii} est la matrice avec tous ses éléments nuls sauf l'élément de la i ème ligne et de la i ème colonne.

Nous obtenons donc que E est invariant si et seulement si (1.13) entraîne (1.12), c'est-à-dire si et seulement si

$$\forall z \in R^{l+p}, \quad z^T T_i z \leq 0, \quad i = 1 \dots p \implies z^T T_0 z \leq 0 \quad (1.14)$$

où

$$T_0 = \begin{pmatrix} A^T P + P A & P B \\ B^T P & 0 \end{pmatrix}, \quad T_i = \begin{pmatrix} -c_i^T c_i & 0 \\ 0 & E_{ii} \end{pmatrix} \quad i = 1 \dots p.$$

En pratique, vérifier ceci pour une matrice P est très difficile. C'est pourquoi les mathématiciens se restreignent à vérifier la condition suivante :

$$\exists \tau_1 \geq 0, \dots, \tau_p \geq 0 \text{ tels que } T_0 \preceq \tau_1 T_1 + \dots + \tau_p T_p$$

qui entraîne indubitablement (1.14) et nous permet de conclure à l'invariance de E . Le problème se résume donc à traiter l'information $T_0 - \tau_1 T_1 - \dots - \tau_p T_p \preceq 0$ faisant référence à la programmation semi-définie.

Ainsi, nous avons défini la programmation semi-définie. Nous utilisons la formulation (1.6) du problème dans les chapitres 2 et 5, relatifs tous les deux à la dualité, et la formulation (1.5) pour les chapitres 3 et 4, relatifs aux méthodes de résolution de ce problème. De plus, par les différentes applications, nous voyons que la programmation semi-définie recouvre une large classe de problèmes. Elle mérite donc notre attention.

Chapitre 2

La dualité

Tout comme en programmation linéaire, il est aisé de vérifier l'admissibilité ou non d'une solution à un problème de programmation semi-définie. Mais, dans le cas admissible, comment vérifier si celle-ci est proche de l'optimalité ? La théorie de la dualité fournit une réponse à cette question. L'idée est d'associer, tout comme en programmation linéaire, un problème de programmation semi-définie dual au problème de départ. Le primal et le dual étant connu, ils nous permettent alors de calculer des bornes pour la valeur optimale recherchée.

Le dual nous donne aussi la possibilité de déterminer l'admissibilité ou non d'un problème. En programmation semi-définie comme en programmation linéaire, la théorie de la dualité s'avère être un outil indispensable.

Puisque la programmation semi-définie est un cas particulier de la programmation conique convexe, nous étudions la dualité dans ce cas plus général et utilisons la formulation (1.6) du problème.

Nous commençons par définir le problème dual d'un problème de programmation semi-définie et les termes nécessaires à la compréhension de la théorie de la dualité en programmation conique convexe. Nous établissons également les propriétés de base des cônes convexes utilisées ultérieurement. Ensuite, nous caractérisons l'admissibilité et l'inadmissibilité d'un problème de programmation conique convexe. Enfin tous ces développements nous permettent de généraliser les résultats de la dualité obtenus en programmation linéaire à la programmation conique convexe. En effet, la relation de dualité faible reste valable en programmation conique convexe mais le théorème prouvant la dualité forte nécessite une hypothèse supplémentaire. Ainsi, les résultats de la dualité en programmation conique convexe sont légèrement plus faibles que ceux de la programmation linéaire.

2.1 Le dual

Soit le problème de programmation conique convexe suivant, $CP(b, c, \mathcal{A}, \mathcal{K})$:

$$\inf c^T x \quad \text{s.c. } x \in (b + \mathcal{A}) \cap \mathcal{K}. \quad (2.1)$$

En programmation semi-définie, nous avons $\mathcal{K} = \mathcal{H}_{\bar{n}}^+$.

Le problème dual associé est

$$\inf b^T z \quad \text{s.c. } z \in (c + \mathcal{A}^\perp) \cap \mathcal{K}^*$$

où, étant donné un cône convexe \mathcal{K} , le cône dual \mathcal{K}^* correspondant est défini comme $\mathcal{K}^* = \{z | z^T \mathcal{K} \subseteq R^+\}$.

Dans le cas de la programmation semi-définie, $\mathcal{K}^* = \mathcal{H}_{\bar{n}}^+$. En effet, la définition de cône dual nous donne

$$(\mathcal{H}_{\bar{n}}^+)^* = \{X \text{ matrice de dimension } \bar{n} | X \bullet Y \geq 0 \text{ où } Y \in \mathcal{H}_{\bar{n}}^+\}.$$

D'une part, nous avons $\mathcal{H}_{\bar{n}}^+ \subseteq (\mathcal{H}_{\bar{n}}^+)^*$. En effet, soit $X \in (\mathcal{H}_{\bar{n}}^+)$. Montrons que $X \in (\mathcal{H}_{\bar{n}}^+)^*$. Soit $Y \in (\mathcal{H}_{\bar{n}}^+)$. Décomposant X et Y , nous avons $X = V_1^T L_1 V_1$ et $Y = V_2^T L_2 V_2$ où L_1 et L_2 sont les matrices diagonales des valeurs propres de X , $l_{1,i}$, et de Y , $l_{2,i}$, et V_1 et V_2 sont des matrices unitaires. Par conséquent, nous avons successivement

$$\begin{aligned} \text{Re tr } XY &= \text{Re tr}(V_1^T L_1 V_2^T L_2 V_2) \\ &= \text{Re tr}(L_1 L_2 V_1^T V_1 V_2^T V_2) \quad \text{voir annexe} \\ &= \text{Re tr}(L_1 L_2) \\ \text{Re tr } XY &= \sum_{i=1}^{\bar{n}} l_{1,i} l_{2,i}. \end{aligned} \tag{2.2}$$

L'hypothèse, X et $Y \in \mathcal{H}_{\bar{n}}^+$, entraîne que $l_{ij} \geq 0$, $i = 1 \dots \bar{n}$, $j = 1, 2$ et dès lors la thèse. D'autre part, si $X \in (\mathcal{H}_{\bar{n}}^+)^*$, nous pouvons montrer que $X \in \mathcal{H}_{\bar{n}}^+$. En effet, par le même raisonnement que celui développé ci-dessus, nous concluons que $X \in (\mathcal{H}_{\bar{n}}^+)^{**}$. Or, comme $\mathcal{H}_{\bar{n}}^+$ est un cône fermé, $(\mathcal{H}_{\bar{n}}^+)^{**} = (\mathcal{H}_{\bar{n}}^+)$ par le théorème 2.1. X est donc un élément de $\mathcal{H}_{\bar{n}}^+$.

Il est facilement observable que le problème dual est aussi un problème de programmation semi-définie, désigné par $CP(c, b, \mathcal{A}^\perp, \mathcal{K}^*)$.

2.2 Terminologie

Après le calcul du dual, définissons les différents termes nécessaires à la compréhension de la théorie de la dualité pour la programmation conique convexe.

Soit un cône convexe \mathcal{K} . Nous définissons

le cône dual : $\mathcal{K}^* := \{z | z^T \mathcal{K} \subseteq R^+\}$. Il est convexe et fermé.

sub $\mathcal{K} := \mathcal{K} \cap -\mathcal{K}$ est le plus grand sous-espace linéaire contenu dans \mathcal{K} .

$\mathcal{K}^\perp := \text{sub} \mathcal{K}^* = \mathcal{K}^* \cap -\mathcal{K}^* = \{z | z^T \mathcal{K} = \{0\}\}$.

span $\mathcal{K} := \mathcal{K} \oplus -\mathcal{K}$ est le sous-espace linéaire engendré par les éléments de \mathcal{K} .

un cône \mathcal{K} pointé si et seulement si $\text{sub } \mathcal{K} := \mathcal{K} \cap -\mathcal{K} = \{0\}$.

un cône \mathcal{K} solide si et seulement si $\text{int } \mathcal{K} \neq \emptyset$.

Soit un problème primal (P) de programmation conique convexe, $CP(b, c, \mathcal{A}, \mathcal{K})$. Nous définissons les termes suivants.

L'ensemble des solutions admissibles de (P) est $\mathcal{F}_P := (b + \mathcal{A}) \cap \mathcal{K}$.

x est une direction si $x \in (\mathcal{A} \cap \mathcal{K})$.

x est une direction intérieure si $x \in (\mathcal{A} \cap \text{rel}\mathcal{K})$ où *rel* désigne l'intérieur relatif.

x est une direction unilatérale si x est une direction et -x n'en est pas une, c'est-à-dire si $x \in ((\mathcal{A} \cap \mathcal{K}) \setminus \text{sub}(\mathcal{A} \cap \mathcal{K}))$.

x est une direction de niveau inférieur si $x \in (\mathcal{A} \cap \mathcal{K})$ et $c^T x \leq 0$.

x est une direction de niveau inférieur unilatérale si x est une direction de niveau inférieur et -x n'en est pas une.

x est une direction améliorante si x est une direction telle que $c^T x < 0$.

$(x_i)_{i \in N}$ est une suite de directions améliorantes si $x_i \in \mathcal{K} \forall i \in N$, $\limsup_{i \rightarrow \infty} c^T x_i < 0$ et $\lim_{i \rightarrow \infty} \text{dist}(x_i, \mathcal{A}) = 0$.

S'il existe une direction améliorante x, en prenant $x_i = x \forall i \in N$, on démontre qu'il existe une suite de directions améliorantes. L'inverse est en général faux.

L'ensemble admissible normalisé de (P) est $\mathcal{F}_P \cap (\text{sub}(\mathcal{A} \cap \mathcal{K}))^\perp$.

L'ensemble des solutions intérieures de (P) est $\text{int}\mathcal{F}_P = \mathcal{F}_P \cap \text{rel}\mathcal{K}$ où *int* désigne l'intérieur.

(P) est admissible si $\mathcal{F}_P \neq \emptyset$.

Par définition de la distance, si (P) est admissible, alors $\text{dist}(b + \mathcal{A}, \mathcal{K}) = 0$.

(P) est fortement admissible si $\text{int}\mathcal{F}_P \neq \emptyset$.

Souvent, cette condition est connue sous le nom de contrainte de qualification de Slater généralisée.

(P) est faiblement admissible si (P) est admissible mais (P) n'est pas fortement admissible.

(P) est faiblement inadmissible si $\text{dist}(b + \mathcal{A}, \mathcal{K}) = 0$ et $\mathcal{F}_P = \emptyset$.

(P) est fortement inadmissible si $\text{dist}(b + \mathcal{A}, \mathcal{K}) > 0$.

La valeur optimale de (P) est $p^* := \inf c^T \mathcal{F}_P$. Nous adoptons la convention suivant laquelle $p^* = \infty$ lorsque (P) est inadmissible.

L'ensemble optimal de (P) est $\mathcal{F}_P^* := \{x \in \mathcal{F}_P \mid c^T x = p^*\}$.

L'ensemble optimal normalisé de (P) est $\mathcal{F}_P^* \cap (\text{sub}(\mathcal{A} \cap \mathcal{K}))^\perp$.

(P) est résolvable si $\mathcal{F}_P^* \neq \emptyset$.

(P) est non borné si $p^* = -\infty$. Notons que si (P) est admissible et s'il existe une direction améliorante, alors (P) est non borné.

Les définitions concernant le dual sont totalement analogues à celles concernant le primal.

Pour clore cette section, définissons le saut de dualité, notion très importante pour les méthodes primales-duales de suivi de chemin.

Si $x \in \mathcal{F}_\mathcal{P}$ et $z \in \mathcal{F}_\mathcal{D}$, alors le saut de dualité de la paire admissible (x, z) est

$$x^T z = b^T z + c^T x \quad (2.3)$$

où $b^T z$ est la valeur de la fonction objectif duale et $c^T x$ est la valeur de la fonction objectif primale. Cette relation découle du fait que $x - b \in \mathcal{A}$ et $z - c \in \mathcal{A}^\perp$ avec $b \in \mathcal{A}^\perp$ et $c \in \mathcal{A}$.

2.3 Propriétés de base des cônes convexes

Nous établissons ici les propriétés des cônes convexes nécessaires aux développements de la théorie de la dualité et des chapitres suivants. Ceux-ci paraîtront quelque peu techniques mais tellement efficaces pour la suite.

Soit le problème primal $CP(b, c, \mathcal{A}, \mathcal{K})$. Pour que le dual du dual $CP(c, b, \mathcal{A}^\perp, \mathcal{K}^*)$ soit équivalent au primal, il suffit que $\mathcal{K} = \mathcal{K}^{**}$. Le théorème 2.1 prouve que ceci n'est vérifié que lorsque \mathcal{K} est fermé.

Théorème 2.1 (Théorème bipolaire) *Soit \mathcal{K} un cône convexe dans \mathbb{R}^n . Alors $cl\mathcal{K} = \mathcal{K}^{**}$.*

Le théorème 2.1, admis, caractérise la fermeture de \mathcal{K} .

Lemme 2.1 *Soient \mathcal{K}_1 et \mathcal{K}_2 deux cônes convexes. Alors $\mathcal{K}_1^* \cap \mathcal{K}_2^* = (\mathcal{K}_1 \oplus \mathcal{K}_2)^*$.*

preuve : Nous avons successivement les équivalences suivantes :

$$\begin{aligned} x \in (\mathcal{K}_1 \oplus \mathcal{K}_2)^* &\iff x^T(\mathcal{K}_1 \oplus \mathcal{K}_2) \subseteq \mathbb{R}^+ \\ &\iff x^T \mathcal{K}_1 \subseteq \mathbb{R}^+ \text{ et } x^T \mathcal{K}_2 \subseteq \mathbb{R}^+ \\ &\iff x \in \mathcal{K}_1^* \cap \mathcal{K}_2^* \end{aligned}$$

cqfd

Corollaire 2.1 *Soit \mathcal{K} un cône convexe, alors $\text{span } \mathcal{K} = (\text{sub}\mathcal{K}^*)^\perp = \mathcal{K}^{\perp\perp}$.*

preuve : Nous avons, d'une part, par définition, $\mathcal{K}^\perp = \text{sub}(\mathcal{K}^*)$. Il suit que $\mathcal{K}^{\perp\perp} = (\text{sub}\mathcal{K}^*)^\perp$.

D'autre part, nous écrivons successivement

$$\begin{aligned} (\text{span}\mathcal{K})^* &= (\mathcal{K} \oplus -\mathcal{K})^* \\ &= \mathcal{K}^* \cap (-\mathcal{K})^* \quad \text{par le lemme 2.1} \\ &= \text{sub}\mathcal{K}^* \end{aligned}$$

Par conséquent, nous avons $(\text{span}\mathcal{K})^{**} = \text{span}\mathcal{K} = (\text{sub}\mathcal{K}^*)^\perp$, puisque $\text{span}\mathcal{K}$ est un sous-espace linéaire.

cqfd

Le corollaire (2.1) indique que \mathcal{K} est solide $\iff \text{span}\mathcal{K} = R^n = (\text{sub}\mathcal{K}^*)^\perp \iff \text{sub}\mathcal{K}^* = 0 \iff \mathcal{K}^*$ est pointé.

Le théorème 2.1 caractérise la fermeture de \mathcal{K} . Le théorème 2.2 caractérise quant à lui l'intérieur relatif du cône \mathcal{K} .

Théorème 2.2 Soit \mathcal{K} un cône convexe dans R^n . Alors

$$x \in \text{rel}\mathcal{K} \iff \begin{cases} x \in \text{span}\mathcal{K} \\ x^T(\mathcal{K}^* \setminus \mathcal{K}^\perp) \subseteq R^{++}. \end{cases}$$

preuve : Supposons d'abord $x \in \text{rel}\mathcal{K}$ et z un vecteur arbitraire non nul de $(\mathcal{K}^* \setminus \mathcal{K}^\perp)$. Montrons que $x^T z > 0$ et $x \in \text{span}\mathcal{K}$.

Soit \hat{z} la projection orthogonale de z sur \mathcal{K} . Par la définition d'intérieur relatif, nous avons qu'il existe $\delta(z) > 0$ tel que $x - \delta(z)\hat{z} \in \mathcal{K}$. Par conséquent, nous écrivons successivement

$$\begin{aligned} 0 &\leq z^T(x - \delta(z)\hat{z}) = z^T x - \delta(z)z^T \hat{z} \\ &= z^T x - \delta(z)\|\hat{z}\|_2^2 < z^T x. \end{aligned}$$

De plus, vu l'inclusion $\text{rel}\mathcal{K} \subseteq \text{span}\mathcal{K}$ évidente, nous concluons $x \in \text{span}\mathcal{K}$.

Supposons ensuite $x \in \text{span}\mathcal{K}$ tel que $x^T(\mathcal{K}^* \setminus \mathcal{K}^\perp) \subseteq R^{++}$ et posons $\epsilon = \inf_z \{x^T z \mid z \in (\mathcal{K}^* \cap \text{span}\mathcal{K}) \text{ et } \|z\| = 1\}$. Par hypothèse $x \in \text{span}\mathcal{K}$ où $\text{span}\mathcal{K}$ est égal à $\mathcal{K}^{\perp\perp}$, par le corollaire 2.1. Il suit que $x^T \mathcal{K}^\perp = \{0\}$. Par définition de \mathcal{K}^\perp et puisque $x^T(\mathcal{K}^* \setminus \mathcal{K}^\perp) \subseteq R^{++}$, nous avons $x^T \mathcal{K}^* \subseteq R^+$, c'est-à-dire que $x \in \mathcal{K}^{**}$. Par conséquent, nous obtenons $x^T z \geq 0 \quad \forall z \in \mathcal{K}^* \cap \text{span}\mathcal{K}$. De plus, pour $z \in \mathcal{K}^* \cap \text{span}\mathcal{K}$ et $z \neq 0$, nous avons $z \in \mathcal{K}^* \setminus \mathcal{K}^\perp$. Donc, nous obtenons $\epsilon > 0$.

Nous avons finalement par construction, $\forall y \in \text{span}\mathcal{K}$, $y \neq 0$ et $z \in \mathcal{K}^*$,

$$z^T(x - \frac{\epsilon}{\|y\|_*} y) = z^T x - \frac{\epsilon}{\|y\|_*} z^T y \geq 0$$

où $\|y\|_* = \max_v \{v^H y \mid \|v\| = 1\}$.

Ceci implique que $x \in \text{rel}\mathcal{K}^{**}$, puisque $\text{span}\mathcal{K} = \text{span}\text{cl}\mathcal{K} = \text{span}\mathcal{K}^{**}$ et $\frac{\epsilon}{\|y\|_*} > 0$. Par le théorème bipolaire, $x \in \text{rel}\mathcal{K}$.

cqfd

Le lemme suivant nous sera par la suite fort utile.

Lemme 2.2 Soient \mathcal{K} un cône convexe dans R^n et $M \in R^{n \times n}$ une matrice inversible. Alors $(M^T \mathcal{K})^* = M^{-1} \mathcal{K}^*$.

preuve : Nous avons

$$\begin{aligned} y \in (M^T \mathcal{K})^* &\iff y^T(M^T \mathcal{K}) \subseteq R^+ \\ &\iff My \in \mathcal{K}^* \\ &\iff y \in M^{-1} \mathcal{K}^*. \end{aligned}$$

cqfd

2.4 Caractérisation de l'admissibilité et de l'inadmissibilité

Dans ce chapitre, nous tentons d'obtenir le plus d'informations possibles sur les valeurs optimales primales et duales, quel que soit le caractère admissible du primal et du dual. Dans cette section, nous caractérisons l'admissibilité et l'inadmissibilité d'un problème de programmation conique convexe. Ces caractérisations nous permettront d'obtenir dans la section suivante les informations souhaitées et nous serons d'une grande utilité dans le chapitre 5.

Les deux théorèmes suivants caractérisent l'admissibilité forte primale.

Théorème 2.3 *Le problème primal (2.1) est fortement admissible si et seulement si il n'existe pas de direction duale de niveau unilatérale.*

preuve : Par définition, (2.1) est fortement admissible

$$\begin{aligned}
 &\iff (b + \mathcal{A}) \cap \text{rel}\mathcal{K} \neq \emptyset \\
 &\iff b \in \mathcal{A} \oplus \text{rel}\mathcal{K} \\
 &\iff b \in \text{rel}\mathcal{A} \oplus \text{rel}\mathcal{K} \quad \text{car } \mathcal{A} \text{ est un sous-espace vectoriel} \\
 &\iff b \in \text{rel}(\mathcal{A} \oplus \mathcal{K}) \\
 &\iff \begin{cases} b \in \text{span}(\mathcal{A} \oplus \mathcal{K}) \\ b^T((\mathcal{A} \oplus \mathcal{K})^* \setminus (\mathcal{A} \oplus \mathcal{K})^\perp) \subseteq R^{++} \end{cases} \quad \text{par le théorème 2.2.}
 \end{aligned}$$

Or, par le corollaire 2.1, nous avons $\text{span}(\mathcal{A} \oplus \mathcal{K}) = (\text{sub}(\mathcal{A} \oplus \mathcal{K})^*)^\perp$, et par le lemme 2.1 et le fait que \mathcal{A} soit un sous-espace vectoriel, nous obtenons $(\mathcal{A} \oplus \mathcal{K})^* = (\mathcal{A}^\perp \cap \mathcal{K}^*)$. De plus, $(\mathcal{A} \oplus \mathcal{K})^\perp$ est par définition égal à $\text{sub}(\mathcal{A} \oplus \mathcal{K})^* = \text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*)$. Par conséquent, nous obtenons l'équivalence suivante : (2.1) est fortement admissible si et seulement si

$$\begin{cases} b^T \text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*) = \{0\} \\ b^T((\mathcal{A}^\perp \cap \mathcal{K}^*) \setminus \text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*)) \subseteq R^{++} \end{cases}$$

Or, z est une direction duale de niveau unilatérale

$$\begin{aligned}
 &\iff \begin{cases} z \in (\mathcal{A}^\perp \cap \mathcal{K}^*) \text{ et } b^T z \leq 0 \\ -z \notin (\mathcal{A}^\perp \cap \mathcal{K}^*) \text{ ou } b^T(-z) > 0 \end{cases} \\
 &\iff \begin{cases} \text{soit } z \in (\mathcal{A}^\perp \cap \mathcal{K}^*) \setminus \text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*) \text{ et } b^T z \leq 0 \\ \text{soit } z \in \text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*) \text{ et } b^T z < 0 \end{cases}
 \end{aligned}$$

Donc, il n'existe pas de direction duale de niveau unilatérale signifie que

$$\begin{cases} \forall z \in (\mathcal{A}^\perp \cap \mathcal{K}^*) \setminus \text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*) \quad b^T z > 0 \\ \forall z \in \text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*) \quad b^T z \geq 0 \end{cases}$$

qui est équivalent à dire

$$\begin{cases} \forall z \in (\mathcal{A}^\perp \cap \mathcal{K}^*) \setminus \text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*) & b^T z > 0 \\ \forall z \in \text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*) & b^T z = 0 \end{cases}$$

Comme développé ci-dessus, ceci est équivalent à dire que (2.1) est fortement admissible.

cqfd

Corollaire 2.2 *Considérant (2.1), nous avons l'équivalence suivante :*

il existe une direction intérieure améliorante primale si et seulement si le dual n'est pas admissible et il n'existe pas de direction duale unilatérale.

preuve : Si nous supposons $c=0$, le dual est admissible car $c \in (\mathcal{A}^\perp \cap \mathcal{K}^*)$. De plus, $\forall x \in (\mathcal{A} \cap \text{rel}\mathcal{K})$, $c^T x = 0$. Ainsi, il n'existe pas de direction intérieure améliorante primale.

Supposons donc $c \neq 0$. Par hypothèse, nous savons que $c \in \mathcal{A}$. De ce fait, il est facile de montrer que

$$-\frac{c}{\|c\|_2^2} + (\mathcal{A} \cap \ker c^T) = \{x \in \mathcal{A} | c^T x = -1\}. \quad (2.4)$$

Par conséquent, x est une direction intérieure primale améliorante pour $CP(b, c, \mathcal{A}, \mathcal{K})$ est équivalent à trouver un $\alpha > 0$ tel que

$$\alpha x \in \left(-\frac{c}{\|c\|_2^2} + (\mathcal{A} \cap \ker c^T)\right) \cap \text{rel}\mathcal{K}.$$

Il suit qu'il existe une direction intérieure primale améliorante pour $CP(b, c, \mathcal{A}, \mathcal{K})$ si et seulement si il existe un point intérieur primal pour $CP(-\frac{c}{\|c\|_2^2}, 0, \mathcal{A} \cap \ker c^T, \mathcal{K})$.

Le théorème 2.3 affirme que ceci est équivalent à ce qu'il n'existe pas de direction duale de niveau unilatérale pour $CP(-\frac{c}{\|c\|_2^2}, 0, \mathcal{A} \cap \ker c^T, \mathcal{K})$.

Montrons que ceci revient à dire que le dual de $CP(b, c, \mathcal{A}, \mathcal{K})$ n'est pas admissible et qu'il n'existe pas de direction duale unilatérale.

En effet, le dual de $CP(-\frac{c}{\|c\|_2^2}, 0, \mathcal{A} \cap \ker c^T, \mathcal{K})$ est $CP(0, -\frac{c}{\|c\|_2^2}, \mathcal{A}^\perp \oplus \text{im } c, \mathcal{K}^*)$ puisque

$$(\mathcal{A} \cap \ker c^T)^\perp = \mathcal{A}^\perp \oplus \text{im } c. \quad (2.5)$$

Par un même raisonnement que celui réalisé dans la preuve du théorème 2.3, nous pouvons dire qu'il n'existe pas de direction duale de niveau unilatérale pour $CP(-\frac{c}{\|c\|_2^2}, 0, \mathcal{A} \cap \ker c^T, \mathcal{K})$ signifie

$$\begin{cases} \forall z \in (((\mathcal{A}^\perp \oplus \text{im } c) \cap \mathcal{K}^*) \setminus \text{sub}((\mathcal{A}^\perp \oplus \text{im } c) \cap \mathcal{K}^*)) & -\frac{c}{\|c\|_2^2}^T z > 0 \\ \forall z \in \text{sub}((\mathcal{A}^\perp \oplus \text{im } c) \cap \mathcal{K}^*) & -\frac{c}{\|c\|_2^2}^T z = 0 \end{cases}$$

Il suit que

$$\forall z \text{ tel que } \exists \alpha \text{ vérifiant } \alpha z \in (c + \mathcal{A}^\perp) \cap \mathcal{K}^*, \quad c^T z \leq 0$$

Soit $c^T z < 0$, dans ce cas, $z \notin (c + \mathcal{A}^\perp)$ puisque $c \in \mathcal{A}$. Par conséquent, le dual de $CP(b, c, \mathcal{A}, \mathcal{K})$ n'est pas admissible.

Soit $c^T z = 0$, dans ce cas $c = 0$ ce qui contredit l'hypothèse. Donc il n'existe pas de $z \in ((\mathcal{A}^\perp \oplus \text{im } c) \cap \mathcal{K}^*)$. Ainsi, il n'existe pas de $z \in (\mathcal{A}^\perp \cap \mathcal{K}^*) \setminus -\mathcal{K}^*$. Par conséquent, $CP(b, c, \mathcal{A}, \mathcal{K})$ n'a pas de directions duales unilatérales.

cqfd

Lemme 2.3 *Considérons $CP(b, c, \mathcal{A}, \mathcal{K})$ fortement admissible. Soit $(z^{(i)})_{i \in N} \in \mathcal{K}^* \cap \text{span}(\mathcal{A} \oplus \mathcal{K})$ telle que*

$$\lim_{i \rightarrow \infty} \text{dist}(z^{(i)}, c + \mathcal{A}^\perp) = 0 \text{ et } \limsup_{i \rightarrow \infty} b^T z^{(i)} < +\infty$$

Alors $(z^{(i)})_{i \in N}$ est une suite bornée.

preuve : Supposons par l'absurde qu'il existe une suite $(z^{(i)})_{i \in N} \in \mathcal{K}^* \cap \text{span}(\mathcal{A} \oplus \mathcal{K})$ telle que

$$\lim_{i \rightarrow \infty} \|z^{(i)}\| = \infty, \lim_{i \rightarrow \infty} \text{dist}(z^{(i)}, c + \mathcal{A}^\perp) = 0 \text{ et } \limsup_{i \rightarrow \infty} b^T z^{(i)} < +\infty.$$

Nous supposons également sans perdre de généralité que $\|z^{(i)}\| > 0 \forall i$ et que $y := \lim_{i \rightarrow \infty} \frac{z^{(i)}}{\|z^{(i)}\|}$ existe. Puisque la suite $(\frac{z^{(i)}}{\|z^{(i)}\|})_{i \in N}$ est contenue dans $\mathcal{K}^* \cap \text{span}(\mathcal{A} \oplus \mathcal{K})$, il suit que

$$y \in \mathcal{K}^* \cap \text{span}(\mathcal{A} \oplus \mathcal{K}) = \mathcal{K}^* \cap (\text{sub}(\mathcal{A} \oplus \mathcal{K})^*)^\perp = \mathcal{K}^* \cap (\text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*))^\perp \quad (2.6)$$

par le corollaire 2.1 et le lemme 2.1. De plus, nous avons par hypothèse

$$y = y - \lim_{i \rightarrow \infty} \frac{1}{\|z^{(i)}\|} c = \lim_{i \rightarrow \infty} \frac{z^{(i)} - c}{\|z^{(i)}\|} \in \mathcal{A}^\perp. \quad (2.7)$$

Ensuite, puisque $\limsup_{i \rightarrow \infty} b^T z^{(i)} < \infty$, nous obtenons

$$b^T y = \lim_{i \rightarrow \infty} \frac{b^T z^{(i)}}{\|z^{(i)}\|} = 0.$$

Par construction, nous avons donc

$$\begin{cases} \|y\| = 1 \\ b^T y = 0 \\ y \in (\mathcal{A}^\perp \cap \mathcal{K}^*) \setminus -\mathcal{K}^* \end{cases}$$

La dernière équation s'obtient par (2.7) pour l'appartenance à \mathcal{A}^\perp , par (2.6) pour l'appartenance à \mathcal{K}^* et par l'absurde pour la non-appartenance à $-\mathcal{K}^*$. En effet, si $y \in -\mathcal{K}^*$, nous avons $\forall k \in \mathcal{K} \quad y^T k = 0$ puisque $y \in \mathcal{K}^*$. Or, $y^T \mathcal{A} = 0 \forall a \in \mathcal{A}$. Par conséquent, $y \in (\mathcal{A} \oplus \mathcal{K})^\perp = \text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*)$. Or, $y \in (\text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*))^\perp$ par (2.6). Ainsi, $y = 0$, ce qui est impossible puisque $\|y\| = 1$. Donc, y est une direction de niveau primale unilatérale. Par le théorème 2.3, $CP(b, c, \mathcal{A}, \mathcal{K})$ n'est pas fortement admissible, ce qui contredit notre hypothèse.

cqfd

Théorème 2.4 *Supposons le dual de $CP(b, c, \mathcal{A}, \mathcal{K})$ admissible.*

$CP(b, c, \mathcal{A}, \mathcal{K})$ est fortement admissible si et seulement si l'ensemble optimal dual normalisé est non vide et borné.

preuve : Supposons $CP(b, c, \mathcal{A}, \mathcal{K})$ fortement admissible. Par le lemme 2.3, nous avons que toute suite $(z^{(i)})_{i \in \mathbb{N}}$ de points duaux admissibles normalisés telle que $\lim_{i \rightarrow \infty} b^T z^{(i)} = d^*$ est une suite bornée. Par conséquent, $c + \mathcal{A}^\perp$ et \mathcal{K}^* étant fermés, nous avons que $\mathcal{F}_\mathcal{D} = (c + \mathcal{A}^\perp) \cap \mathcal{K}^*$ est fermé et différent du vide car $\lim_{i \rightarrow \infty} z^{(i)}$ y appartient. Rappelons que l'ensemble optimal dual normalisé est

$$\{z \in \mathcal{F}_\mathcal{D} \cap (\text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*))^\perp \mid b^T z = d^*\}$$

Cet ensemble est différent du vide car $\lim_{i \rightarrow \infty} z^{(i)}$ y appartient. Il est également borné. En effet, s'il ne l'était pas, il existerait $(g^{(i)})_{i \in \mathbb{N}}$ une suite de cet ensemble non bornée. Or, par le lemme 2.3, toute suite de cet ensemble est bornée.

Supposons que l'ensemble optimal dual normalisé est non vide et borné. Alors il n'existe pas de directions de niveau duale unilatérale. Sinon,

$$\begin{cases} \exists z \in (\mathcal{A}^\perp \cap \mathcal{K}^*) \setminus \text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*) \text{ et } b^T z \leq 0 \\ \text{ou } z \in \text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*) \text{ et } b^T z < 0 \end{cases}$$

Si $\exists z \in (\mathcal{A}^\perp \cap \mathcal{K}^*)$ tel que $b^T z < 0$, puisque le dual est admissible, le dual n'a pas de solution optimale bornée, ce qui contredit notre hypothèse.

Si $\exists z \in (\mathcal{A}^\perp \cap \mathcal{K}^*) \setminus \text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*)$ et $b^T z = 0$, alors, puisque $b \in \mathcal{A}^\perp$, nous avons $z \in \mathcal{A}$. Ceci entraîne que z soit nul, ce qui est impossible puisque $z \notin \text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*)$.

Par le théorème 2.3, $CP(b, c, \mathcal{A}, \mathcal{K})$ est fortement admissible.

cqfd

Corollaire 2.3 *Supposons le dual de $CP(b, c, \mathcal{A}, \mathcal{K})$ admissible.*

L'ensemble dual admissible normalisé est borné et non vide si et seulement si il existe une direction intérieure primale.

preuve : Considérons $N = CP(0, c, \mathcal{A}, \mathcal{K})$. Nous avons par le théorème 2.4 appliqué à ce nouveau problème,

$\text{int}\mathcal{F}_\mathcal{N} = (\mathcal{A} \cap \text{rel}\mathcal{K}) \neq \emptyset \iff \{z \in \mathcal{F}_\mathcal{D} \cap (\text{sub}(\mathcal{A}^\perp \cap \mathcal{K}^*))^\perp\}$ est différent du vide et borné qui est la thèse.

cqfd

Caractérisons ensuite l'inadmissibilité par les deux lemmes suivants du type Farkas.

Lemme 2.4 (Premier lemme de type Farkas) *Le primal $CP(b, c, \mathcal{A}, \mathcal{K})$ est fortement inadmissible si et seulement si il existe une direction duale améliorante.*

preuve : Raisonnons par la contraposée. Par définition, le primal $CP(b, c, \mathcal{A}, \mathcal{K})$ n'est pas fortement inadmissible

$$\begin{aligned}
&\Longleftrightarrow \text{dist}(b + \mathcal{A}, \mathcal{K}) = 0 \\
&\Longleftrightarrow \exists (x^{(i)})_{i \in N} \in \mathcal{K} \text{ telle que } \lim_{i \rightarrow \infty} \mathcal{P}_{\mathcal{A}^\perp} x^{(i)} = b \\
&\Longleftrightarrow b \in \text{cl} \mathcal{P}_{\mathcal{A}^\perp} \mathcal{K} \\
&\Longleftrightarrow b \in \mathcal{P}_{\mathcal{A}^\perp} \mathcal{K}^{**} \\
&\text{par le théorème 2.1 et le fait qu'un cône convexe projeté reste un cône convexe,} \\
&\Longleftrightarrow b^T \mathcal{P}_{\mathcal{A}^\perp} \mathcal{K}^* \subseteq R^+.
\end{aligned}$$

Or,

$$\begin{aligned}
\mathcal{P}_{\mathcal{A}^\perp} \mathcal{K}^* &= \{s \in R^n \mid s^T \mathcal{P}_{\mathcal{A}^\perp} \mathcal{K} \subseteq R^+\} \\
&= \{s \in R^n \mid \mathcal{P}_{\mathcal{A}^\perp} s \in \mathcal{K}^*\} \\
&\text{car une matrice de projection est symétrique} \\
&= (\mathcal{K}^* \cap \mathcal{A}^\perp) + \mathcal{A}.
\end{aligned}$$

Tenant compte de ce développement, il suit que le primal $CP(b, c, \mathcal{A}, \mathcal{K})$ n'est pas fortement inadmissible

$$\begin{aligned}
&\Longleftrightarrow b^T ((\mathcal{K}^* \cap \mathcal{A}^\perp) + \mathcal{A}) \subseteq R^+ \\
&\Longleftrightarrow b^T (\mathcal{K}^* \cap \mathcal{A}^\perp) \subseteq R^+ \quad \text{car } b \in \mathcal{A}^\perp \\
&\Longleftrightarrow \text{il n'existe pas de directions duales améliorantes.}
\end{aligned}$$

cqfd

Lemme 2.5 (Second lemme de type Farkas) *Le dual de $CP(b, c, \mathcal{A}, \mathcal{K})$ est admissible si et seulement si il n'existe pas de suites de directions améliorantes primales.*

preuve : Cette preuve est très semblable à celle du corollaire 2.2.

Considérons le cas où $c=0$. Le dual de $CP(b, c, \mathcal{A}, \mathcal{K})$ est alors admissible, $0 \in \mathcal{F}_\mathcal{D}$ et il n'existe pas de suites de directions améliorantes primales puisque $c^T x = 0, \forall x \in \mathcal{A} \cap \mathcal{K}$.

Ensuite, considérons le cas où c est différent de 0. Nous avons

$$c^T z > 0 \text{ et } z \in (\mathcal{A}^\perp \oplus \text{im } c) \cap \mathcal{K}^* \Longleftrightarrow \exists \alpha > 0 \text{ tel que } \alpha z \in (c + \mathcal{A}^\perp) \cap \mathcal{K}^*.$$

Cette équivalence est obtenue en décomposant z en $a + cl$ où $l \in R$ et $a \in \mathcal{A}^\perp$. La condition $c^T z > 0$ impose à α d'être strictement positif. Tenant compte de cette équivalence, nous obtenons que le dual de $CP(b, c, \mathcal{A}, \mathcal{K})$ est admissible

$$\begin{aligned}
&\Longleftrightarrow (c + \mathcal{A}^\perp) \cap \mathcal{K}^* \neq \emptyset \\
&\Longleftrightarrow CP(-\frac{c}{\|c\|_2^2}, 0, (\mathcal{A}^\perp \oplus \text{im } c)^\perp, \mathcal{K}) \text{ a une direction duale améliorante} \\
&\Longleftrightarrow \text{dist}(-\frac{c}{\|c\|_2^2} + (\mathcal{A}^\perp \oplus \text{im } c)^\perp, \mathcal{K}) > 0 \text{ par le lemme 2.4.}
\end{aligned}$$

Or, par la preuve du corollaire 2.2, (2.4) et (2.5), nous avons

$$-\frac{c}{\|c\|_2^2} + (\mathcal{A}^\perp \oplus \text{im } c)^\perp = -\frac{c}{\|c\|_2^2} + (\mathcal{A} \cap \ker c^T) = \{x \in \mathcal{A} | c^T x = -1\}.$$

Par conséquent, le dual de $CP(b, c, \mathcal{A}, \mathcal{K})$ est admissible si et seulement si il n'existe pas de suites de directions améliorantes primales.

cqfd

Nous résumons dans le tableau suivant, les équivalences caractérisant l'admissibilité primale de problème de programmation conique convexe fermée sous l'hypothèse d'admissibilité duale. Puisque la dualité est tout à fait symétrique pour les problèmes de ce type, nous pouvons établir un tableau analogue caractérisant l'admissibilité duale.

Primal	Dual
1. Direction intérieure	Ensemble admissible normalisé borné et non vide
2. Fortement admissible	Ensemble optimal normalisé borné et non vide
3. Faiblement admissible	Direction de niveau unilatérale mais pas de suite de directions améliorantes
4. Faiblement inadmissible	Suite de directions améliorantes mais pas de directions améliorantes
5. Fortement inadmissible	Direction améliorante
6. Fortement inadmissible et pas de direction unilatérale	Direction intérieure améliorante

TAB. 2.1: équivalences caractérisant le caractère admissible de problèmes de programmation conique convexe primal fermé dont le dual est admissible.

Toutes les caractérisations mises en évidence dans ce tableau sont des applications directes de ce qui précède. En effet, la première ligne du tableau 2.1 est le corollaire 2.3, la seconde représente le théorème 2.4, et nécessite l'hypothèse d'admissibilité duale. La troisième découle de la contraposée du théorème 2.3 et du lemme 2.5 appliqué au dual. Ainsi, nous devons supposer \mathcal{K} fermé par le théorème bipolaire, afin que le dual du dual soit le primal. La quatrième ligne suit de la contraposée des lemmes 2.4 et 2.5. La cinquième ligne est la transcription du lemme 2.4. Finalement, la sixième ligne provient du corollaire 2.2 appliqué au dual et du lemme 2.4.

2.5 Dualité faible et dualité forte

Le but de ce chapitre est, comme souligné plus haut, d'obtenir des informations sur la valeur optimale primale d'un problème de programmation conique convexe. Si nous examinons les résultats de la dualité en programmation linéaire, nous avons une relation de dualité faible, $p^* \geq d^*$, et un théorème de dualité forte, les valeurs optimales du primal et du dual sont toujours égales sauf lorsque les deux problèmes sont simultanément inadmissibles. La dualité en programmation linéaire est donc très forte. Notre objectif est à présent de généraliser les résultats obtenus en programmation linéaire à la programmation conique convexe.

Nous démontrons d'abord deux lemmes. Ceux-ci nous conduiront à un théorème menant à la relation de la dualité faible en programmation conique convexe. Le théorème de dualité forte en programmation conique convexe est prouvé ensuite. Cependant, il nécessite une hypothèse plus forte que pour la programmation linéaire.

Lemme 2.6 Soit $\gamma \in R$. Considérons $CP(b, c, \mathcal{A}, \mathcal{K})$ tel que $\text{dist}(b + \mathcal{A}, \mathcal{K}) = 0$.

Alors

$$\begin{pmatrix} b \\ \gamma \end{pmatrix} ((\mathcal{K}^* \times R^+) \cap M_c^{-1}(\mathcal{A}^\perp \times R)) \subseteq R^+ \iff d^* \geq -\gamma$$

$$\text{avec } M_c := \begin{pmatrix} I_{n \times n} & -c \\ 0 & 1 \end{pmatrix}.$$

preuve : Considérons $CP(b, c, \mathcal{A}, \mathcal{K})$ tel que $\text{dist}(b + \mathcal{A}, \mathcal{K}) = 0$.

Raisonnons par la contraposée et supposons $d^* < -\gamma$. Par définition de d^* , nous avons qu'il existe $z \in \mathcal{F}_\mathcal{D}$ tel que $b^T z + \gamma < 0$. Considérons le vecteur $\begin{pmatrix} z \\ 1 \end{pmatrix}$. Nous voyons que

$$\begin{pmatrix} z \\ 1 \end{pmatrix} \in ((\mathcal{K}^* \times R^+) \cap M_c^{-1}(\mathcal{A}^\perp \times R))$$

$$\text{et que } (b^T, \gamma) \begin{pmatrix} z \\ 1 \end{pmatrix} = b^T z + \gamma < 0.$$

Pour démontrer l'autre implication, raisonnons également par la contraposée et supposons qu'il existe

$$\begin{pmatrix} z \\ z_{n+1} \end{pmatrix} \in ((\mathcal{K}^* \times R^+) \cap M_c^{-1}(\mathcal{A}^\perp \times R)) \text{ tel que } b^T z + z_{n+1} \gamma < 0.$$

Notons que si $z_{n+1} = 0$, alors z est une direction duale améliorante, ce qui contredit l'hypothèse $\text{dist}(b + \mathcal{A}, \mathcal{K}) = 0$, par le tableau 2.1. Donc, nous obtenons $z_{n+1} > 0$ et nous avons $\frac{z}{z_{n+1}} \in \mathcal{F}_\mathcal{D}$ puisque $z \in \mathcal{F}_\mathcal{D}$. Nous obtenons finalement $d^* \leq b^T \frac{z}{z_{n+1}} < -\gamma$.

cqfd

Lemme 2.7 Soit $\gamma \in R$. Considérons $CP(b, c, \mathcal{A}, \mathcal{K})$. Alors, nous avons

$$\text{dist}\left(\begin{pmatrix} b \\ \gamma \end{pmatrix} + M_c^T(\mathcal{A} \times \{0\}), \mathcal{K} \times R^+\right) = 0 \iff \bar{p} \leq \gamma$$

où $M_c := \begin{pmatrix} I_{n \times n} & -c \\ 0 & 1 \end{pmatrix}$, et $\bar{p} := \lim_{\epsilon \rightarrow 0} \inf_x \{c^T x \mid x \in \mathcal{K} \text{ dist}(x, b + \mathcal{A}) < \epsilon\}$, la sous-valeur du primal.

preuve : Nous savons par définition que $\bar{p} \leq \gamma$ si et seulement si

il existe une suite $(x^{(i)})_{i \in N} \in \mathcal{K}$ telle que $\lim_{i \rightarrow \infty} \text{dist}(x^{(i)}, b + \mathcal{A}) = 0$ et $\lim_{i \rightarrow \infty} c^T x^{(i)} \leq \gamma$. (2.8)

Supposons $\bar{p} \leq \gamma$ et posons $x_{n+1}^{(i)} := \max(0, \gamma - c^T x^{(i)})$, $i \in N$. Nous obtenons une nouvelle suite $(x^{(i)}, x_{n+1}^{(i)})_{i \in N} \in \mathcal{K} \times R^+$ et telle que

$$\lim_{i \rightarrow \infty} \text{dist}\left(\begin{pmatrix} x^{(i)} \\ x_{n+1}^{(i)} \end{pmatrix}, \begin{pmatrix} b \\ \gamma \end{pmatrix} + M_c^T(\mathcal{A} \times \{0\})\right) = 0.$$

En effet, d'une part, nous avons par hypothèse que $\lim_{i \rightarrow \infty} \text{dist}(x^{(i)}, b + \mathcal{A}) = 0$.

D'autre part, montrons que $\lim_{i \rightarrow \infty} \text{dist}_{a \in \mathcal{A}}(\max(0, \gamma - c^T x^{(i)}), \gamma - c^T a) = 0$. Nous avons successivement

$$\begin{aligned} & \lim_{i \rightarrow \infty} \inf_{a \in \mathcal{A}} |\max(0, \gamma - c^T x^{(i)}) - (\gamma - c^T a)| \\ &= \lim_{i \rightarrow \infty} \inf_{a \in \mathcal{A}} |\max(0, \gamma - c^T x^{(i)}) - (\gamma - c^T(b + a))| \\ &\leq \lim_{i \rightarrow \infty} |\max(0, \gamma - c^T x^{(i)}) - (\gamma - c^T(b + a^{(i)}))| \end{aligned}$$

où $a^{(i)} \in \mathcal{A} \forall i$ et $|x^{(i)} - (b + a^{(i)})|$ tend vers 0 lorsque i tend vers l'infini.

$$\leq \lim_{i \rightarrow \infty} |\max(0, \gamma - c^T x^{(i)}) - (\gamma - c^T x^{(i)})|.$$

Posons $\lim_{i \rightarrow \infty} \gamma - c^T x^{(i)} = \bar{\gamma}$ où $\bar{\gamma} \geq 0$. Nous disposons de deux cas possibles. Soit il y a un nombre fini de $i \in N$ pour lesquels $\gamma - c^T x^{(i)} < 0$, dans ce cas, $\lim_{i \rightarrow \infty} \max(0, \gamma - c^T x^{(i)}) = \bar{\gamma}$. Soit il y a un nombre infini de $i \in N$ pour lesquels $\gamma - c^T x^{(i)} < 0$, dans ce cas, $\lim_{i \rightarrow \infty} \gamma - c^T x^{(i)} = \bar{\gamma} = 0$ et $\lim_{i \rightarrow \infty} \max(0, \gamma - c^T x^{(i)}) = \bar{\gamma} = 0$. Ce développement nous permet de conclure que

$$\text{dist}\left(\begin{pmatrix} b \\ \gamma \end{pmatrix} + M_c^T(\mathcal{A} \times \{0\}), \mathcal{K} \times R^+\right) = 0.$$

De manière inverse, s'il existe une suite $(x^{(i)}, x_{n+1}^{(i)})_{i \in N} \in \mathcal{K} \times R^+$ telle que

$$\lim_{i \rightarrow \infty} \text{dist}\left(\begin{pmatrix} x^{(i)} \\ x_{n+1}^{(i)} \end{pmatrix}, \begin{pmatrix} b \\ \gamma \end{pmatrix} + M_c^T(\mathcal{A} \times \{0\})\right) = 0,$$

alors $(x^{(i)})_{i \in N}$ satisfait (2.8).

cqfd

Le résultat suivant est une relation entre la sous-valeur primale et la valeur optimale primale. Celui-ci conduit à la relation de dualité faible.

Théorème 2.5 *Considérons $CP(b, c, \mathcal{A}, \mathcal{K})$. Si le dual de $CP(b, c, \mathcal{A}, \mathcal{K})$ est inadmissible et le primal est fortement inadmissible, alors $\bar{p} = d^* = \infty$. Dans les autres cas, $\bar{p} = -d^*$.*

preuve : Supposons l'inadmissibilité duale et la forte inadmissibilité primale. Sous la seconde hypothèse, nous avons $\text{dist}(b + \mathcal{A}, \mathcal{K}) > 0$. Dans ce cas, $\bar{p} = \infty$. Or, $d^* = \infty$ par hypothèse. Par conséquent, $\bar{p} = d^* = \infty$.

Supposons, ensuite, l'admissibilité duale et la forte inadmissibilité primale. Dans ce cas, la seconde hypothèse entraîne $\bar{p} = \infty$. Or, il suit du lemme 2.4 qu'il existe une direction améliorante duale. Par conséquent, $d^* = -\infty = -\bar{p}$.

Terminons en supposant que le primal n'est pas fortement inadmissible. Nous avons dans ce cas $\text{dist}(b + \mathcal{A}, \mathcal{K}) = 0$. Nous savons par le lemme 2.7 que $\bar{p} \leq \gamma$

$$\begin{aligned} \iff & \text{dist}\left(\begin{pmatrix} b \\ \gamma \end{pmatrix} + M_c^T(\mathcal{A} \times \{0\}), \mathcal{K} \times R^+\right) = 0 \\ \iff & \begin{pmatrix} b \\ \gamma \end{pmatrix} [(M_c^T(\mathcal{A} \times \{0\}))^\perp \cap (\mathcal{K} \times R^+)^*] \subseteq R^+ \\ & \text{par la contraposée du lemme 2.4} \\ \iff & \begin{pmatrix} b \\ \gamma \end{pmatrix} [M_c^{-1}(\mathcal{A}^\perp \times R) \cap (\mathcal{K}^* \times R^+)] \subseteq R^+ \end{aligned}$$

par le lemme 2.2.

Le lemme 2.6 nous permet de dire, puisque $\text{dist}(b + \mathcal{A}, \mathcal{K}) = 0$, que $\bar{p} \leq \gamma \iff d^* \geq -\gamma \quad \forall \gamma \in R$. Par conséquent, $\bar{p} = -d^*$.

cqfd

Du théorème 2.5 ainsi que de l'inégalité $p^* \geq \bar{p}$ par définition de \bar{p} , suit la relation de dualité faible :

$$p^* \geq -d^*. \quad (2.9)$$

Nous montrons à présent que sous la condition de Slater généralisée, $\text{int}\mathcal{F}_P \neq \emptyset$, la sous-valeur primale coïncide avec la valeur optimale primale, $\bar{p} = p^*$. Il suit alors, par le théorème précédent, la généralisation du théorème de la dualité forte de la programmation linéaire à la programmation semi-définie, sous la condition de Slater généralisée.

Théorème 2.6 (Dualité forte de Slater) *Supposons $\text{int}\mathcal{F}_P \neq \emptyset$. Alors $p^* = \bar{p} = -d^*$. De plus, si $p^* > -\infty$, alors $\mathcal{F}_D^* \neq \emptyset$ et l'ensemble optimal dual normalisé est borné.*

preuve : Supposons sans perdre de généralité \mathcal{K} fermé. Ainsi, nous avons par le théorème 2.1, $\mathcal{K} = \mathcal{K}^{**}$. Définissons la sous-valeur duale \bar{d} comme

$$\bar{d} := \liminf_{\epsilon \rightarrow 0} \inf_z \{b^T z \mid z \in \mathcal{K}^* \text{ dist}(z, c + \mathcal{A}^\perp) < \epsilon\}.$$

Nous appliquons le théorème 2.5 à $CP(c, b, \mathcal{A}^\perp, \mathcal{K}^*)$ puisque $\text{int}\mathcal{F}_\mathcal{P} \neq \emptyset$ et nous obtenons $p^* = -\bar{d} \geq -d^*$ par (2.9). Nous concluons déjà que si $p^* = -\infty$, alors $d^* = \infty$, et le théorème est vérifié.

Il reste donc à considérer le cas où $p^* = -\bar{d} > -\infty$. Par définition de \bar{d} , cette condition entraîne qu'il existe une suite $(z^{(i)})_{i \in \mathbb{N}} \in \mathcal{K}^* \cap \text{span}(\mathcal{A} \oplus \mathcal{K})$ telle que

$$\lim_{i \rightarrow \infty} \text{dist}(z^{(i)}, c + \mathcal{A}^\perp) = 0 \quad \text{et} \quad \limsup_{i \rightarrow \infty} b^T z^{(i)} = -p^* < +\infty$$

Comme le primal est fortement admissible, nous appliquons le lemme 2.3, et obtenons que $(z^{(i)})_{i \in \mathbb{N}}$ est une suite bornée. Celle-ci possède donc un point d'adhérence $z^{(\infty)}$. Evidemment, $z^{(\infty)} \in \mathcal{F}_\mathcal{D}$, $b^T z^{(\infty)} = -p^*$. Par conséquent, nous avons $-p^* \leq d^* \leq b^T z^{(\infty)} = -p^*$. Ainsi, nous obtenons $p^* = -d^*$ et $\mathcal{F}_\mathcal{D}^* \neq \emptyset$.

Le caractère non borné de l'ensemble des solutions optimales duales normalisé suit du théorème 2.4.

cqfd

Le théorème 2.6 implique que si \mathcal{K} est fermé et $\text{int}\mathcal{F}_\mathcal{P} \times \text{int}\mathcal{F}_\mathcal{D} \neq \emptyset$, alors $\mathcal{F}_\mathcal{P}^* \times \mathcal{F}_\mathcal{D}^* \neq \emptyset$, ainsi que

$$\begin{aligned} (x^{*T} z^*) &= c^T x^* + b^T z^* \quad \forall (x^*, z^*) \in \mathcal{F}_\mathcal{P}^* \times \mathcal{F}_\mathcal{D}^* \text{ par (2.3)} \\ &= p^* + d^* \\ &= 0. \end{aligned}$$

Suite à cette considération, nous définissons une solution complémentaire pour un problème de programmation conique convexe comme étant une paire $(x, z) \in \mathcal{F}_\mathcal{P} \times \mathcal{F}_\mathcal{D}$ telle que $x^T z = 0$. Evidemment, toute paire de solution complémentaire est une paire de solution optimale. L'inverse cependant est en général faux, à l'exception du cas où le primal et le dual sont fortement admissibles (cfr théorème 2.6.). En fait, la dualité forte est nécessaire mais pas suffisante pour l'existence d'une solution complémentaire.

Terminons par résumer dans le tableau ci-dessous, les relations de dualité pour la programmation conique convexe $CP(b, c, \mathcal{A}, \mathcal{K})$ où \mathcal{K} est un cône fermé. Tous les éléments de ce tableau découlent des théorèmes 2.5 et 2.6. Nous désignons par (P) , le primal $CP(b, c, \mathcal{A}, \mathcal{K})$ et par (D) , son problème dual.

Notons les éléments de la première ligne de ce tableau A_1, A_2, A_3, A_4 , ceux de la seconde B_1, B_2, B_3, B_4 , ceux de la troisième C_1, C_2, C_3, C_4 et ceux de la quatrième D_1, D_2, D_3 et D_4 .

En ce qui concerne A_1 , le primal et le dual sont fortement admissibles. Par le théorème 2.6,

			PRIMAL			
			ADMISSIBLE		INADMISSIBLE	
			FORT	FAIBLE	FAIBLE	FORT
DUAL	ADMIS-SIBLE	FORT	$p^* = -d^*$, (P) et (D) résolvables	$p^* = -d^*$, (P) résolvable	(D) non borné, $p^* = \infty$	(D) non borné, $p^* = \infty$
		FAIBLE	$p^* = -d^*$, (D) résolvable	$p^* \geq -d^*$	$p^* \geq -d^*$, $p^* = \infty$	(D) non borné, $p^* = \infty$
	INADMIS-SIBLE	FAIBLE	(P) non borné, $d^* = \infty$	$p^* \geq -d^*$ $d^* = \infty$	$p^* = d^* =$ ∞	$p^* = d^* =$ ∞
		FORT	(P) non borné, $d^* = \infty$	(P) non borné, $d^* = \infty$	$p^* = d^* =$ ∞	$p^* = d^* =$ ∞

TAB. 2.2: relations de dualité en programmation conique convexe fermée

$p^* = -d^*$. De plus, p^* est différent de $-\infty$, sinon le dual ne serait pas admissible. De même, d^* est différent de $-\infty$. Par conséquent, par le théorème 2.6, (P) et (D) sont résolubles. A_2 et B_1 sont obtenus par un raisonnement similaire. C_1 et D_1 suivent du théorème 2.6 appliqué au primal tandis que A_3 et A_4 suivent du théorème 2.6 appliqué au dual. B_4 et D_2 proviennent du lemme 2.4. Les éléments restants sont dûs à la relation de faible dualité (2.9) et ont une valeur optimale infinie dans les cas inadmissibles. Mentionnons que la symétrie du tableau est due au fait que nous considérons la dualité pour la programmation conique convexe fermée.

Chapitre 3

Méthodes de suivi de chemin en programmation semi-définie

Les chapitres 1 et 2 nous ont permis de cerner le problème de la programmation semi-définie. Dans ce chapitre, nous allons présenter une méthode de résolution de ce problème en le considérant comme cas particulier de la programmation conique convexe.

La méthode choisie consiste d'une part en une méthode primale-duale. En effet, vu les liens étroits explicités dans le chapitre 2 entre un primal et son dual ainsi qu'entre les valeurs optimales respectives, la méthode qui s'intéresse à résoudre ces deux problèmes simultanément se révèle très efficace du point de vue numérique. D'autre part, la méthode dégagée est aussi une méthode de points intérieurs. En effet, engendrer des itérés à l'intérieur du domaine admissible plutôt que sur sa frontière permet généralement une progression plus rapide vers une solution optimale. Nous supposons donc, dans ce chapitre, l'admissibilité forte primale et duale.

De façon plus précise, nous voulons élaborer un algorithme général où les itérés du type $(X^{(k)}, Z^{(k)})$ où $X^{(k)} \in \text{int}\mathcal{F}_\mathcal{P}$ et $Z^{(k)} \in \text{int}\mathcal{F}_\mathcal{D}$ convergent vers (X^*, Z^*) où X^* est une solution optimale du primal et Z^* celle du dual. En d'autres mots, à partir d'un point de départ $(X^{(0)}, Z^{(0)})$, nous devons construire des points successifs $(X^{(k)}, Z^{(k)})$ sur base du calcul d'une direction et d'une longueur de pas sans perdre de vue l'idée de la convergence de ces itérés vers une solution optimale (X^*, Z^*) .

Notre manière de procéder est la suivante. Nous choisirons tout d'abord dans la section 3.1. une transformation qui transportera l'espace $\mathcal{F}_\mathcal{P} \times \mathcal{F}_\mathcal{D}$ dans un nouvel espace rendant nos variables matricielles (X, Z) plus aisément manipulables. Ceci facilitera l'obtention de la direction de Newton. La direction de descente sera d'abord calculée dans l'espace dit transformé pour être ensuite reconvertie dans l'espace initial. Cet aller-retour entre l'espace de départ $\mathcal{F}_\mathcal{P} \times \mathcal{F}_\mathcal{D}$ et l'espace transformé, réalisé dans la section 3.2., imposera aux itérés $(X^{(k)}, Z^{(k)})$ de suivre approximativement une trajectoire spécifique. La méthode obtenue est donc une méthode de suivi de chemin. Celle-ci, exigeant une certaine proximité de l'itéré vis-à-vis de la trajectoire, conduira à la définition de voisinage (section 3.3.). Dans les sections 3.4. et 3.5, nous bâtirons un algorithme général de résolution

d'un problème de programmation semi-définie et de son dual par une méthode de points intérieurs primale-duale et de suivi de chemin. Celui-ci nous permettra de généraliser trois algorithmes de la programmation linéaire et de les analyser. Vu les nombreuses études réalisées pour la résolution de problème de programmation semi-définie par une méthode de points intérieurs primale-duale de suivi de chemin, la section 3.6 rappelle quelques directions connues et effectue le lien avec la direction de Newton de la section 3.2.

3.1 Le V-espace

La méthode présentée ici étant une méthode primale-duale, nous commençons par rassembler le problème de programmation semi-définie primal et son dual en un seul problème appelé problème primal-dual. L'objet de cette section est alors de choisir une transformation qui transportera nos variables de l'espace initial $\mathcal{F}_P \times \mathcal{F}_D$, nommées variables initiales, dans un nouvel espace présentant plusieurs avantages tels la simplicité des variables ou des propriétés facilitant le calcul de la direction de Newton.

Notre recherche de la matrice de transformation se réalise en deux temps. Tout d'abord, après avoir examiné un éventail de possibilités de celle-ci pour des valeurs fixées de (X, Z) (3.1.1), nous porterons dans 3.1.2. notre préférence pour une matrice de transformation symétrique primale-duale L_d . En effet, celle-ci transforme X et Z en une seule même matrice diagonale V . D'où l'appellation du "V-espace". Les avantages de cette transformation analysés dans 3.1.3. et le calcul rapide de L_d et de V dans 3.1.4. confirmeront notre choix. Nous devrons également généraliser cette transformation au cas où les variables (X, Z) n'ont plus chacune une valeur fixée mais deviennent dépendantes d'un paramètre de longueur t , c'est-à-dire lorsque $(X(t), Z(t))$ décrit une trajectoire primale $X(t)$ et une duale $Z(t)$ (3.1.5.).

3.1.1 Transformation du problème initial

Commençons par rassembler le problème de programmation semi-définie primal et son dual en un seul problème primal-dual.

Rappelons qu'un problème de programmation semi-définie peut s'écrire de la manière suivante (1.4) :

$$(P) : \inf\{C \bullet X \mid X \in (B + \mathcal{A}) \cap \mathcal{H}_n^+\} \text{ où } B, C, X \in \mathcal{H}_n^+ \text{ et } \mathcal{A} \subseteq \mathcal{H}_n.$$

Son dual se note (voir section 2.1.) :

$$(D) : \inf\{B \bullet Z \mid Z \in (C + \mathcal{A}^\perp) \cap \mathcal{H}_n^+\} \text{ où } B, C, Z \in \mathcal{H}_n^+ \text{ et } \mathcal{A}^\perp \subseteq \mathcal{H}_n.$$

Résoudre le primal et le dual en même temps revient à rechercher la solution du problème de programmation semi-définie suivant, appelé problème primal-dual :

$$\inf\{C \bullet X + B \bullet Z \mid X \in \mathcal{F}_P \text{ et } Z \in \mathcal{F}_D\}.$$

Or, faisant appel au résultat (2.3) du chapitre 2, nous voyons que $C \bullet X + B \bullet Z = X \bullet Z$ est le saut de dualité du problème primal-dual, $C \bullet X$ étant la fonction objectif primale à minimiser et $B \bullet Z$ son analogue dual. De plus, (2.9) affirme que la somme $C \bullet X + B \bullet Z$ est toujours positive.

Ensuite, l'hypothèse d'admissibilité forte primale et duale garantit, de nouveau par le chapitre 2, l'existence d'une solution complémentaire (X, Z) . Cette solution est par définition telle que $(X, Z) \in \mathcal{F}_P \times \mathcal{F}_D$ et $X \bullet Z = 0$.

Par conséquent, sous l'hypothèse de forte admissibilité primale et duale, résoudre (P) et (D) équivaut à résoudre :

$$\min\{X \bullet Z \mid X \in \mathcal{F}_P \text{ et } Z \in \mathcal{F}_D\} \quad (3.1)$$

dont la valeur optimale est nulle et réalisée par une solution dite complémentaire.

Comme remarqué plus haut, le calcul de la direction à prendre en un point (X, Z) , lorsque X et Z sont des matrices, relève d'une assez grande difficulté. C'est pourquoi nous envisageons de transporter le problème (3.1) dans un nouvel espace dit espace transformé. Ainsi, nous pourrions travailler avec des variables plus simples, comme par exemple des matrices diagonales.

L'objectif de ce point est donc d'effectuer une transformation des problèmes primal et dual initiaux de manière générale afin d'obtenir un nouveau problème primal-dual équivalent à (3.1). Dans le point 3.1.3., grâce à certains critères de sélection, nous nous restreindrons à une transformation particulière garantissant des variables matricielles diagonales.

Pour cela, nous nous basons sur une propriété bien connue du cône semi-défini positif :

$$\forall L \in C^{\bar{n} \times \bar{n}} \text{ inversible} : L^{-1}XL^{-H} \in \mathcal{H}_{\bar{n}}^+ \iff X \in \mathcal{H}_{\bar{n}}^+.$$

Cette relation se note de manière plus concise : $(L^{-1} \otimes_H L^{-1})\mathcal{H}_{\bar{n}}^+ = \mathcal{H}_{\bar{n}}^+$ (cfr annexe) et indique que la transformation linéaire $(L^{-1} \otimes_H L^{-1})$ qui, à une matrice X carrée de dimension \bar{n} , fait correspondre $L^{-1}XL^{-H}$ n'affecte pas le cône $\mathcal{H}_{\bar{n}}^+$. Par ailleurs, elle transforme le sous-espace linéaire \mathcal{A} en un sous-espace linéaire différent noté

$$\begin{aligned} \mathcal{A}(L) &:= (L^{-1} \otimes_H L^{-1})\mathcal{A} \\ &= \{L^{-1}XL^{-H} \mid X \in \mathcal{A}\} \\ &= \{Y \mid LY L^H \in \mathcal{A}\}. \end{aligned}$$

A l'aide de cette transformation, nous obtenons un problème équivalent au problème de départ (P),

$$(P_L) : \inf\{(L^HCL) \bullet \bar{X} \mid \bar{X} \in (L^{-1}BL^{-H} + \mathcal{A}(L)) \cap \mathcal{H}_{\bar{n}}^+\}$$

où $\bar{X} = L^{-1}XL^{-H}$.

Remarquons que (P_L) est un problème de programmation conique convexe, noté

$$CP(L^{-1}BL^{-H}, L^HCL, \mathcal{A}(L), \mathcal{H}_n^+).$$

Pour montrer l'équivalence entre (P) et (P_L) , c'est-à-dire entre $CP(B, C, \mathcal{A}, \mathcal{H}_n^+)$ et $CP(L^{-1}BL^{-H}, L^HCL, \mathcal{A}(L), \mathcal{H}_n^+)$, il suffit d'observer que les fonctions objectifs et les domaines admissibles des deux problèmes sont les mêmes. En effet, la fonction objectif de $CP(B, C, \mathcal{A}, \mathcal{H}_n^+)$ peut s'écrire successivement

$$\begin{aligned} C \bullet X &= \operatorname{Re} \operatorname{tr} X^H C \\ &= \operatorname{Re} \operatorname{tr} L^{-1} X^H C L \text{ (voir annexe)} \\ &= \operatorname{Re} \operatorname{tr} L^{-1} X^H L^{-H} L^H C L \\ &= L^H C L \bullet L^{-1} X L^{-H} \\ &= L^H C L \bullet \bar{X}, \end{aligned}$$

qui est la fonction objectif de $CP(L^{-1}BL^{-H}, L^HCL, \mathcal{A}(L), \mathcal{H}_n^+)$, et la contrainte de $CP(B, C, \mathcal{A}, \mathcal{H}_n^+)$ est

$$\begin{aligned} X \in (B + \mathcal{A}) \cap \mathcal{H}_n^+ &\Leftrightarrow L^{-1}XL^{-H} \in (L^{-1}BL^{-H} + (L^{-1} \otimes_H L^{-1})\mathcal{A}) \cap (L^{-1} \otimes_H L^{-1})\mathcal{H}_n^+ \\ &\Leftrightarrow \bar{X} \in (L^{-1}BL^{-H} + \mathcal{A}(L)) \cap \mathcal{H}_n^+ \end{aligned}$$

qui est la contrainte de $CP(L^{-1}BL^{-H}, L^HCL, \mathcal{A}(L), \mathcal{H}_n^+)$.

Associons ensuite au problème primal (P_L) son dual (D_L)

$$CP(L^HCL, L^{-1}BL^{-H}, \mathcal{A}^\perp(L), \mathcal{H}_n^+),$$

si l'on se réfère à la section 2.1. Celui-ci se note donc

$$(D_L) : \inf \{ (L^{-1}BL^{-H}) \bullet \bar{Z} \mid \bar{Z} \in (L^HCL + \mathcal{A}^\perp(L)) \cap \mathcal{H}_n^+ \}.$$

Remarquons que $\mathcal{A}^\perp(L)$ peut s'écrire sous la forme

$$\mathcal{A}^\perp(L) = (L^H \otimes_H L^H) \mathcal{A}^\perp = \{ \bar{Z} \mid L^{-H} \bar{Z} L^{-1} \in \mathcal{A}^\perp \}$$

En effet, nous avons successivement

$$\begin{aligned} \mathcal{A}^\perp(L) &= \operatorname{sub}(\mathcal{A}(L)^*) \text{ par définition de } \perp \\ &= \mathcal{A}(L)^* \text{ car } \mathcal{A}(L) \text{ est un sous espace linéaire} \\ &= (L^{-1} \mathcal{A} L^{-H})^* \text{ par définition de } \mathcal{A}(L) \\ &= L^H (\mathcal{A} L^{-H})^* \text{ par le lemme 2.2} \\ &= L^H \mathcal{A}^\perp L \text{ par le lemme 2.2} \\ &= (L^H \otimes_H L^H) \mathcal{A}^\perp \text{ voir annexe} \\ &= \{ L^H Z L \mid Z \in \mathcal{A}^\perp \} \\ &= \{ \bar{Z} \mid L^{-H} \bar{Z} L^{-1} \in \mathcal{A}^\perp \}. \end{aligned}$$

Naturellement, nous pouvons également définir les ensembles admissibles des problèmes (P_L) et (D_L) .

$$\begin{aligned}\mathcal{F}_P(L) &:= (L^{-1}BL^{-H} + \mathcal{A}(L)) \cap \mathcal{H}_n^+ \\ \mathcal{F}_D(L) &:= (L^HCL + \mathcal{A}^\perp(L)) \cap \mathcal{H}_n^+.\end{aligned}$$

Notons également que si $L=I$, (P_L) est alors identiquement (P) ainsi que (D_L) est (D) .

Avec les notations décrites ci-dessus, le problème primal-dual associé aux problèmes (P_L) et (D_L) s'écrit

$$\min\{\bar{X} \bullet \bar{Z} \mid \bar{X} \in \mathcal{F}_P(L) \text{ et } \bar{Z} \in \mathcal{F}_D(L)\}, \quad (3.2)$$

en effectuant un raisonnement analogue à celui réalisé sur les problèmes de départ (P) et (D) au début de ce point. Nous pouvons dire également que sous l'hypothèse de forte admissibilité primale et duale, résoudre (P_L) et (D_L) revient donc à un problème (3.2) de minimisation du saut de dualité dont la valeur optimale est nulle et réalisée par une solution complémentaire.

Tout le développement précédent est justifié par le fait que les problèmes (3.1) et (3.2) sont équivalents pour toute matrice L inversible. En effet, nous remarquons tout d'abord l'invariance du saut de dualité $X \bullet Z$ sous la transformation $(L^{-1} \otimes_H L^{-1})$.

En effet, $\bar{X} = L^{-1}XL^{-H}$ et $\bar{Z} = L^HZL$. Donc,

$$\begin{aligned}\bar{X} \bullet \bar{Z} &= \operatorname{Re} \operatorname{tr} \bar{Z}^H \bar{X} \\ &= \operatorname{Re} \operatorname{tr} L^H Z^H L L^{-1} X L^{-H} \\ &= \operatorname{Re} \operatorname{tr} L^H Z^H X L^{-H} \\ &= \operatorname{Re} \operatorname{tr} Z^H X \quad \text{voir annexe} \\ &= X \bullet Z\end{aligned}$$

De plus, $\bar{X} \in \mathcal{F}_P(L) \Leftrightarrow X \in \mathcal{F}_P$ et $\bar{Z} \in \mathcal{F}_D(L) \Leftrightarrow Z \in \mathcal{F}_D$.

Ainsi pour résoudre un problème de programmation semi-définie de la forme (3.1) ainsi que son dual, il suffit de résoudre le problème de minimisation (3.2) correspondant, où L est une matrice inversible.

3.1.2 Choix de la matrice de transformation L pour un point intérieur primal-dual (X, Z)

Par le point précédent, nous connaissons la condition que L , matrice de transformation, doit satisfaire impérativement : être inversible. Néanmoins, il reste encore un vaste champ de possibilités pour le choix de cette matrice de transformation.

Afin d'atténuer la difficulté de ce choix, nous nous centrons, dans cette section, sur un point intérieur primal-dual et admissible du problème (3.1), soit (X, Z) , et nous nous demandons quelle valeur de L donne le problème transformé le plus avantageux.

Une première idée, intéressante du point de vue primal, consiste à prendre $L = X^{\frac{1}{2}}$. En effet, dans ce cas, nous avons

$$\bar{X} = L^{-1} X L^{-H} = X^{-\frac{1}{2}} X (X^{-\frac{1}{2}})^H = X^{\frac{1}{2}} (X^{-\frac{1}{2}})^H = I$$

puisque $X \in \mathcal{H}$.

Ce choix de $L = X^{\frac{1}{2}}$ a la propriété non-négligeable de centrer \bar{X} . Celle-ci permet d'éviter le risque de non-convergence des itérés dû aux petits pas réalisés si X se situe à proximité de la frontière du domaine admissible primal. Cette propriété est traduite en mathématique de la manière suivante :

$$\bar{X} + \{\Delta \bar{X} \in \mathcal{H} \mid \|\Delta \bar{X}\|_F < 1\} \subset \mathcal{H}_{\bar{n}}^{++}. \quad (3.3)$$

En effet, soit un léger accroissement $\Delta \bar{X}$ de \bar{X} tel que $\|\Delta \bar{X}\|_F < 1$.

Montrons que $\bar{X} + \Delta \bar{X} \in \mathcal{H}_{\bar{n}}^{++}$, c'est-à-dire que, étant donné $y \in C^{\bar{n}}$ et $y \neq 0$, nous avons $y^H(\bar{X} + \Delta \bar{X})y > 0$, et que $\bar{X} + \Delta \bar{X} \in \mathcal{H}_{\bar{n}}$.

Tout d'abord, nous avons $y^H(\bar{X} + \Delta \bar{X})y = y^H \bar{X} y + y^H \Delta \bar{X} y$ où $y^H \bar{X} y = \|y\|_2^2 > 0$ et

$$\begin{aligned} |y^H \Delta \bar{X} y| &\leq \|y^H \Delta \bar{X}\|_2 \|y\|_2 \\ &\leq \|y^H \Delta \bar{X}\|_F \|y\|_2 \text{ voir annexe} \\ &\leq \|y^H\|_2 \|\Delta \bar{X}\|_F \|y\|_2 \text{ voir annexe} \\ &< \|y^H\|_2 \|y\|_2 \text{ car } \|\Delta \bar{X}\|_F < 1 \\ &= \|y\|_2^2 \end{aligned}$$

et donc $-\|y\|_2 < y^H \Delta \bar{X} y < \|y\|_2$. Par conséquent, nous affirmons que $0 < y^H(\bar{X} + \Delta \bar{X})y$. Ensuite, nous avons $\bar{X} + \Delta \bar{X} \in \mathcal{H}_{\bar{n}}$ puisque $\bar{X} = I \in \mathcal{H}_{\bar{n}}$ et $\Delta \bar{X} \in \mathcal{H}_{\bar{n}}$ par hypothèse. nous concluons que $\bar{X} + \Delta \bar{X} \in \mathcal{H}_{\bar{n}}^{++}$.

Cependant, ce beau résultat ne s'applique pas nécessairement à la solution duale transformée $\bar{Z} = X^{\frac{1}{2}} Z X^{\frac{1}{2}}$. \bar{Z} peut donc se trouver près de la frontière de $\mathcal{H}_{\bar{n}}^{++}$.

De plus, cette transformation $L = X^{\frac{1}{2}}$ ne nous permet pas de voir comment réduire le saut de dualité $\bar{X} \bullet \bar{Z} = \text{Re tr } X^{\frac{1}{2}} Z X^{\frac{1}{2}}$, L dépendant directement de notre variable primale X . Ne pouvant pallier à ces deux inconvénients, nous abandonnons notre premier choix.

Puisque nous ne désirons pas une matrice de transformation L qui se base de manière explicite sur le problème primal ou sur le problème dual, la seconde idée pousse à définir L comme une matrice de transformation symétrique primale-duale. Une telle matrice apporte, comme nous allons le voir par la suite, l'avantage considérable de se ramener, à

partir de nos deux variables, X et Z , matrices hermitiennes, à une seule variable, V , matrice diagonale à éléments positifs.

Tout d'abord, définissons une matrice de transformation symétrique primale-duale : L_d est une matrice de transformation symétrique primale-duale pour (X, Z) si et seulement si $L_d^{-1} X L_d^{-H} = L_d^H Z L_d$.
Remarquons que pour ceci L_d doit être inversible.

Ensuite, caractérisons une telle matrice de transformation L_d par le lemme suivant :

Lemme 3.1 *Etant donné X et $Z \in \mathcal{H}_n^{++}$, une matrice $\bar{n} \times \bar{n}$ inversible L_d est une transformation symétrique primale-duale pour X et Z , c'est-à-dire par définition que $L_d^{-1} X L_d^{-H} = L_d^H Z L_d$ si et seulement si $L_d L_d^H = D(X, Z)$ où $D(X, Z) := Z^{-\frac{1}{2}} (Z^{\frac{1}{2}} X Z^{\frac{1}{2}})^{\frac{1}{2}} Z^{-\frac{1}{2}}$.*

preuve : Par définition, une matrice L_d de transformation symétrique primale-duale pour (X, Z) satisfait

$$\begin{aligned}
L_d^{-1} X L_d^{-H} = L_d^H Z L_d &\iff X L_d^{-H} = L_d L_d^H Z L_d \\
&\iff X = L_d L_d^H Z L_d L_d^H \\
&\iff X = (Z^{-\frac{1}{2}} Z^{\frac{1}{2}}) L_d L_d^H Z L_d L_d^H (Z^{\frac{1}{2}} Z^{-\frac{1}{2}}) \\
&\iff X = Z^{-\frac{1}{2}} (Z^{\frac{1}{2}} L_d L_d^H Z^{\frac{1}{2}})^2 Z^{-\frac{1}{2}} \\
&\iff Z^{\frac{1}{2}} X Z^{\frac{1}{2}} = (Z^{\frac{1}{2}} L_d L_d^H Z^{\frac{1}{2}})^2 \\
&\iff (Z^{\frac{1}{2}} X Z^{\frac{1}{2}})^{\frac{1}{2}} = Z^{\frac{1}{2}} L_d L_d^H Z^{\frac{1}{2}} \\
&\iff Z^{-\frac{1}{2}} (Z^{\frac{1}{2}} X Z^{\frac{1}{2}})^{\frac{1}{2}} Z^{-\frac{1}{2}} = L_d L_d^H
\end{aligned}$$

cqfd

Evidemment, pour X et Z fixés appartenant à \mathcal{H}_n^{++} , L_d n'est pas définie de manière unique. En effet, pour calculer L_d , nous pouvons utiliser la factorisation de Cholesky de $D(X, Z)$. Dans ce cas, L_d est triangulaire inférieure. Nous pouvons aussi prendre la racine carrée de $D(X, Z)$. Dans ce cas, L_d est hermitienne ($D(X, Z) \in \mathcal{H}$).

Mais nous voyons que si L_d est une transformation symétrique primale-duale, alors $L_d Q$ où Q est une matrice unitaire d'ordre \bar{n} est aussi une matrice de transformation symétrique primale-duale. En effet, considérant

- Q tel que $Q Q^H = I = Q^H Q$
- L_d tel que $L_d^{-1} X L_d^{-H} = L_d^H Z L_d$ où X et $Z \in \mathcal{H}_n^{++}$ fixées

vérifions que

$$(L_d Q)^{-1} X (L_d Q)^{-H} = (L_d Q)^H Z (L_d Q).$$

Nous avons successivement

$$\begin{aligned}
Q^{-1} L_d^{-1} X (Q^{-1} L_d^{-1})^H &= Q^H (L_d)^H Z L_d Q \\
L_d^{-1} X L_d^{-H} &= L_d^H Z L_d
\end{aligned}$$

qui est satisfait.

cqfd

De plus, il est possible de choisir une matrice unitaire d'ordre \bar{n} , Q , de telle manière que $(L_d Q)^H Z (L_d Q) = Q^H (L_d^H Z L_d) Q$ soit une matrice diagonale à éléments positifs puisque le point Z transformé valant $L_d^H Z L_d$ est hermitien défini positif ($Z \in \mathcal{H}_{\bar{n}}^{++}$).

Nous pouvons donc conclure qu'il existe une transformation symétrique primale-duale L_d qui transforme les solutions primale et duale, X et Z , en une matrice diagonale à éléments positifs, V , où $V = L_d^{-1} X L_d^{-H} = L_d^H Z L_d$. V est une matrice du V -espace et l'espace transformé est V -espace \times V -espace.

Poussant le raisonnement un peu plus loin, nous remarquons que

$$XZ = L_d (L_d^{-1} X L_d^{-H}) (L_d^H Z L_d) L_d^{-1} = L_d V^2 L_d^{-1}, \quad (3.4)$$

ce qui implique que les éléments diagonaux de la matrice V^2 sont exactement les valeurs propres de la matrice XZ .

Nous avons donc trouvé une matrice de transformation $L = L_d$ correspondant à une solution (X, Z) telle que $\bar{X} = \bar{Z}$ soit une matrice diagonale positive. De plus, cette matrice L_d maintient d'une certaine façon l'avantage de notre premier choix $L_d = X^{\frac{1}{2}}$ qui était de centrer la solution primale X et pallier ses deux inconvénients, l'un faisant référence à la solution duale non nécessairement centrée et l'autre au saut de dualité. Nous développons ceci dans le point suivant.

3.1.3 Avantages de la matrice de transformation L_d et trajectoire centrale

Le premier avantage à souligner est que, contrairement à la transformation primale $L = X^{\frac{1}{2}}$, la transformation symétrique primale-duale L_d ne fait pas de discrimination entre le primal et le dual mais transforme les deux variables de départ X et Z en une seule variable diagonale V . Les matrices diagonales étant beaucoup plus faciles à traiter, nous travaillerons par la suite en grande partie sur V , c'est-à-dire dans le V -espace.

Le second avantage à mettre en évidence est que, similairement à la transformation primale ($L = X^{\frac{1}{2}}$), la transformation symétrique primale-duale centre les solutions X et Z mais malheureusement cette fois-ci à une condition : celle de se trouver sur la trajectoire centrale.

En effet, définissons la trajectoire centrale primale-duale comme étant

$$\text{CPATH} := \{(X, Z) \in \mathcal{F}_{\mathcal{P}} \times \mathcal{F}_{\mathcal{D}} \mid XZ = \mu I \quad \text{où} \quad \mu = \frac{\text{Retr} XZ}{\bar{n}}\}.$$

Similairement à (3.3), nous avons

$$V + \{\Delta V \in \mathcal{H} \mid \|\Delta V\|_F < \mu\} \subset \mathcal{H}_{\bar{n}}^{++}. \quad (3.5)$$

En effet, soit un léger accroissement ΔV de V tel que $\|\Delta V\|_F < \mu$.

Montrons que $V + \Delta V \in \mathcal{H}_{\bar{n}}^{++}$, c'est-à-dire que, étant donné $y \in C^{\bar{n}}$ et $y \neq 0$, nous avons

$y^H(V + \Delta V)y > 0$ et $V + \Delta V \in \mathcal{H}_{\bar{n}}$.

Tout d'abord, nous avons $y^H(V + \Delta V)y = y^H V y + y^H \Delta V y$ où $y^H V y = \sqrt{\mu} \|y\|_2^2 > 0$. En effet, $(X, Z) \in \text{CPATH}$ c'est-à-dire que $XZ = \mu I = L_d V^2 L_d^{-1} \iff L_d^{-1} \mu L_d = V^2 \iff \mu I = V^2 \iff \sqrt{\mu} I = V$;

De plus, nous écrivons successivement

$$\begin{aligned} |y^H \Delta V y| &\leq \|y^H \Delta V\|_2 \|y\|_2 \\ &\leq \|y^H \Delta V\|_F \|y\|_2 \text{ voir annexe} \\ &\leq \|y^H\|_2 \|\Delta V\|_F \|y\|_2 \text{ voir annexe} \\ &< \sqrt{\mu} \|y^H\|_2^2 \text{ car } \|\Delta V\|_F < \sqrt{\mu} \end{aligned}$$

et donc, il suit que $-\sqrt{\mu} \|y\|_2 < y^H \Delta V y < \sqrt{\mu} \|y\|_2$. Par conséquent, nous pouvons affirmer $0 < y^H(V + \Delta V)y$.

Ensuite, nous avons $V + \Delta V \in \mathcal{H}_{\bar{n}}$ puisque V est diagonale et $\Delta V \in \mathcal{H}_{\bar{n}}$ par hypothèse. Nous concluons que $V + \Delta V \in \mathcal{H}_{\bar{n}}^{++}$.

De par ce résultat, nous pouvons affirmer que si l'itéré (X, Z) se trouve sur la trajectoire centrale, nous pouvons prendre un grand pas dans l'espace primal transformé ainsi que dans l'espace dual transformé. En effet, l'itéré résultant $V + \Delta V$ du V -espace restera défini positif. Ainsi, puisque $(L^{-1} \otimes_H L^{-1}) \mathcal{H}_{\bar{n}}^{++} = \mathcal{H}_{\bar{n}}^{++}$, le nouvel itéré de l'espace initial $(X + \Delta X, Z + \Delta Z)$ sera également défini positif. C'est pourquoi, il nous paraît nécessaire dans nos futurs algorithmes d'avoir nos itérés, X et Z , très proches de la trajectoire centrale.

Remarquons que celle-ci est invariante sous la transformation L_d car $XZ = \mu I \iff \bar{X} \bar{Z} = L_d^{-1} X L_d^{-H} L_d^H Z L_d = L_d^{-1} X Z L_d = L_d^{-1} \mu I L_d = \mu I$. Et donc suivre la trajectoire centrale dans l'espace initial ou dans l'espace transformé est équivalent.

Un troisième avantage est que la transformation symétrique primale-duale nous permet de savoir comment réduire le saut de dualité $X \bullet Z$. En effet, si (X, Z) se trouve sur la trajectoire centrale alors $X \bullet Z = \text{Re tr } XZ = \bar{n} \mu$. Donc, si nous imposons à nos itérés de se trouver sur la trajectoire centrale et si nous nous débrouillons pour faire tendre μ vers zéro, alors nos itérés convergeront vers (X, Z) où $X \bullet Z = 0$, qui est bien la valeur optimale de notre problème (3.1).

En conclusion, dans les différents algorithmes présentés dans ce mémoire, la direction sera obtenue via le V -espace. De plus, nous imposerons à nos itérés $(X^{(k)}, Z^{(k)})$ de se trouver proches de la trajectoire centrale, et nous ferons tendre μ vers zéro. Nous n'exigerons pas de ceux-ci de suivre exactement la trajectoire centrale, le calcul des itérés s'avérerait trop coûteux. Maintenant, nous sommes à même de comprendre le nom de "méthodes de suivi de chemin".

3.1.4 Calcul de L_d et de V

Voici la procédure permettant de calculer numériquement L_d et V , toutes deux matrices carrées $\bar{n} \times \bar{n}$.

Soient X et $Z \in \mathcal{H}_{\bar{n}}^{++}$.

Algorithme 3.1 (Transformation symétrique primale-duale)

1. Calculer le facteur de Cholesky L_X de $X \in \mathcal{H}_{\bar{n}}^{++}$, c'est-à-dire calculer L_X tel que $X = L_X L_X^H$
2. Calculer la décomposition spectrale (Q, Λ_{XZ}) de $L_X^H Z L_X \in \mathcal{H}_{\bar{n}}^{++}$, c'est-à-dire calculer Q et Λ_{XZ} tels que $L_X^H Z L_X = Q^H \Lambda_{XZ} Q$
3. Poser $L_d = L_X Q^H \Lambda_{XZ}^{-\frac{1}{4}}$ et $V = \Lambda_{XZ}^{\frac{1}{2}}$

En effet, nous écrivons successivement

$$\begin{aligned}
 XZ & \stackrel{(1)}{=} L_X L_X^H Z \\
 & = L_X L_X^H Z L_X L_X^{-1} \\
 & \stackrel{(2)}{=} L_X Q^H \Lambda_{XZ} Q L_X^{-1} \\
 & = L_X Q^H \Lambda_{XZ}^{-\frac{1}{4}} (\Lambda_{XZ}^{\frac{1}{2}})^2 \Lambda_{XZ}^{\frac{1}{4}} Q L_X^{-1} \\
 & \stackrel{(3)}{=} L_d V^2 L_d^{-1} \quad \text{ce qui confirme 3.4}
 \end{aligned}$$

La procédure ci-dessus consiste en une factorisation de Cholesky d'une matrice hermitienne $\bar{n} \times \bar{n}$ et en une décomposition spectrale d'une matrice hermitienne $\bar{n} \times \bar{n}$. Ces deux décompositions impliquent $O(\bar{n}^3)$ opérations. (voir [3])

3.1.5 Choix de la matrice de transformation $G(t)$ de points intérieurs primaux-duaux $(X(t), Z(t))$

Jusqu'à présent, nous avons calculé grâce à une transformation symétrique primale-duale, L_d , le point V du V -espace correspondant à un point intérieur primal-dual (X, Z) . Dans ce paragraphe-ci, nous nous intéressons aux points $V(t)$ du V -espace correspondants aux points intérieurs primaux-duaux $(X(t), Z(t))$, c'est-à-dire lorsque $X(t)$ décrit une trajectoire admissible intérieure primale et $Z(t)$ son analogue dual.

Considérons un point intérieur (X, Z) . Par définition, nous avons $X \in (B + \mathcal{A}) \cap \mathcal{H}_{\bar{n}}^{++}$ et $Z \in (C + \mathcal{A}^\perp) \cap \mathcal{H}_{\bar{n}}^{++}$. Soit V la matrice diagonale du V -espace correspondante. En d'autres mots, $V = L_d^{-1} X L_d^{-H} = L_d^H Z L_d$ où V est une matrice diagonale à éléments positifs et L_d est la matrice de transformation symétrique primale-duale développée dans le point précédent. Supposons que $(X(t), Z(t))$ définit une trajectoire régulière débutant en $(X=X(0), Z=Z(0))$ dans un voisinage de $(X=X(0), Z=Z(0))$ et ne comportant que des paires $(X(t), Z(t))$ primales-duales intérieures admissibles. t est un paramètre de longueur.

Si nous appliquons la transformation L_d à cette trajectoire, nous obtenons évidemment :

$$\bar{X}(t) := L_d^{-1} X(t) L_d^{-H} \text{ et } \bar{Z}(t) := L_d^H Z(t) L_d.$$

Notons que $\bar{X}(0) = L_d^{-1} X L_d^{-H} = L_d^H Z L_d = \bar{Z}(0) = V$.

Cependant, rien ne nous laisse penser que $\bar{X}(t) = \bar{Z}(t)$ pour $t \neq 0$. Pour cette raison, nous allons appliquer une seconde transformation symétrique primale-duale à $(X(t), Z(t))$ en effectuant le même raisonnement sur $(\bar{X}(t), \bar{Z}(t))$ pour obtenir une matrice $V(t) = \bar{X}(t) = \bar{Z}(t)$ que celui réalisé précédemment uniquement sur (X, Z) pour obtenir $V = \bar{X} = \bar{Z}$.

Nous définissons donc $G(t)$ comme vérifiant

$$G(t)G(t)^H = D(\bar{X}(t), \bar{Z}(t))$$

c'est-à-dire que $G(t)$ est une transformation symétrique primale-duale (voir lemme 3.1.).

Par définition, celle-ci vérifie $V(t) := G(t)^H \bar{Z}(t) G(t) = G(t)^{-1} \bar{X}(t) G(t)^{-H}$.

Or, nous avons $V(0) = V = G(0)^H \bar{Z}(0) G(0) = G(0)^{-1} \bar{X}(0) G(0)^{-H}$ et $V = \bar{Z}(0) = \bar{X}(0)$. Nous devons donc imposer $G(0) = I$.

Par conséquent, $V(t)$ est la trajectoire du V -espace correspondant à $(X(t), Z(t))$, calculée en utilisant la transformation symétrique primale-duale $L_d G(t)$.

Remarquons cependant que les conditions $G(t)G(t)^H = D(\bar{X}(t), \bar{Z}(t))$ et $G(0) = I$ ne définissent pas $G(t)$ de manière unique. Précédemment, L_d définie seulement par le lemme 3.1 ne l'était pas non plus. De plus, outre ces deux conditions, aucune restriction n'est faite sur $G(t)$ et rien ne nous permet de dire que $V(t)$ est diagonale. Par contre, nous avons choisi L_d notre matrice de transformation symétrique primale-duale de notre espace initial vers l'espace transformé comme étant la multiplication d'une matrice d'une transformation symétrique primale-duale par une matrice unitaire pour obtenir une matrice V diagonale. Pour la suite, nous choisissons une transformation spécifique en posant $G(t) = D(\bar{X}(t), \bar{Z}(t))^{\frac{1}{2}}$. De cette manière, $G(t)$ est hermitienne pour tout t , vu que $D(\bar{X}(t), \bar{Z}(t)) \in \mathcal{H}$. Nous commenterons davantage ce choix dans la section 3.6.

Les liens entre le problème initial et le problème transformé étant précisés, nous sommes à même de calculer la direction à prendre en un point. Cette recherche s'effectuera dans l'espace transformé par souci de facilité et sera convertie ensuite dans l'espace initial.

3.2 Calcul de la direction de Newton en un point intérieur primal-dual

Comme mentionné dans 3.1., nous voulons suivre de manière approximative la trajectoire centrale. En d'autres mots, à chaque itération, nous désirons à partir d'un point intérieur $(X(0), Z(0))$ effectuer un pas Newton relativement au système suivant :

$$\begin{cases} X(t)Z(t) = \mu I \\ X(t) \in (B + \mathcal{A}) \cap \mathcal{H}_n^{++} \\ Z(t) \in (C + \mathcal{A}^\perp) \cap \mathcal{H}_n^{++}. \end{cases}$$

Ainsi, après avoir calculé la direction de Newton en ayant pris un point de la trajectoire centrale pour cible et après avoir choisi une longueur de pas t^* , nous obtiendrons un nouvel itéré $(X(t^*), Z(t^*))$, à partir duquel nous pourrions recommencer une nouvelle itération mais en prenant une valeur plus petite pour μ , c'est-à-dire une autre cible sur la trajectoire centrale plus proche d'une solution optimale.

L'objectif de cette section est le calcul de la direction de descente en un point intérieur primal-dual. La recherche de celle-ci étant relativement plus aisée dans l'espace transformé, nous la tirons d'abord de cet espace et ensuite la ramenons dans l'espace initial. Remarquons que, dans cette section, nous ne nous intéressons point à la longueur de pas optimale à chaque itération, t^* . C'est pourquoi nous posons ici $t^*=1$.

3.2.1 Calcul de la direction de Newton en un point intérieur primal-dual $(\bar{X}(0), \bar{Z}(0))$ de l'espace transformé

Soient

$X(0) \in (B + \mathcal{A}) \cap \mathcal{H}_{\bar{n}}^{++}$ un point intérieur de l'espace primal initial,

$Z(0) \in (C + \mathcal{A}^\perp) \cap \mathcal{H}_{\bar{n}}^{++}$ un point intérieur de l'espace dual initial,

$\bar{X}(0) = L_d^{-1} X L_d^{-H}$ le point primal correspondant à $X(0)$ dans l'espace transformé,

$\bar{Z}(0) = L_d^H Z L_d$ le point dual correspondant à $Z(0)$ dans l'espace transformé.

Rappelons que le système correspondant à l'espace initial et à partir duquel nous voulons calculer la direction de Newton est :

$$\begin{cases} X(t)Z(t) = \mu I \\ X(t) \in (B + \mathcal{A}) \cap \mathcal{H}_{\bar{n}}^{++} \\ Z(t) \in (C + \mathcal{A}^\perp) \cap \mathcal{H}_{\bar{n}}^{++}. \end{cases}$$

Le système analogue correspondant cette fois à l'espace transformé est donc :

$$\begin{cases} \bar{X}(t)\bar{Z}(t) = \mu I \\ \bar{X}(t) \in (V + \mathcal{A}(L_d)) \\ \bar{Z}(t) \in (V + \mathcal{A}^\perp(L_d)) \end{cases}$$

en tenant compte du fait que

$$\bar{X}(t) \in (L_d^{-1} B L_d^{-H} + \mathcal{A}(L_d)) \cap \mathcal{H}_{\bar{n}}^{++},$$

$$\bar{Z}(t) \in (L_d^H C L_d + \mathcal{A}^\perp(L_d)) \cap \mathcal{H}_{\bar{n}}^{++},$$

$$V = \bar{X}(0) = \bar{Z}(0).$$

De plus, nous n'imposons pas à $\bar{X}(t)$ et à $\bar{Z}(t)$ d'être définies positives maintenant. Nous nous arrangerons pour que ces deux matrices le soient par la suite en vérifiant notre proximité de la trajectoire centrale.

Nous pouvons encore améliorer quelque peu ce système. En effet, comme cité plus haut, nous posons $t^*=1$ dans cette section. Ensuite, $X(t)Z(t) = V^2(t)$. Ceci est donc vrai aussi pour $t^*=1$. De plus, nous nous intéressons seulement aux trajectoires $X(t)$ et $Z(t)$ affines en t . C'est pourquoi, nous avons

$$\begin{aligned} X(t) &= X(0) + tX^{(1)}(0) = V + tX^{(1)}(0) = V + tD_X \\ Z(t) &= Z(0) + tZ^{(1)}(0) = V + tZ^{(1)}(0) = V + tD_Z \end{aligned} \quad (3.6)$$

où $D_X := X^{(1)}(0) \in \mathcal{A}(L_d)$ et $D_Z := Z^{(1)}(0) \in \mathcal{A}^\perp(L_d)$.

Ceci est donc vrai également pour $t=1$. Finalement, nous introduisons comme dans le cas de la programmation linéaire un paramètre de centrage $\gamma \in [0,1]$. Notre saut de dualité devient par conséquent $X(t) \bullet Z(t) = X(t) \bullet Z(t) = \bar{n}\gamma\mu$, et notre équation du système $X(t)Z(t) = \gamma\mu I$ dans l'espace initial, ou $X(t)Z(t) = \gamma\mu I$ dans l'espace transformé. Nous voyons donc qu'en prenant $\gamma=1$, nous visons un point de la trajectoire centrale et que, par contre, en posant $\gamma=0$, nous visons la solution optimale. Choisir $0 < \gamma < 1$ concilie ces deux objectifs.

Ainsi, le système à partir duquel nous allons calculer la direction de Newton (D_X, D_Z) de l'espace transformé est :

$$\begin{cases} V(1)^2 = \gamma\mu I \\ D_X \in \mathcal{A}(L_d) \\ D_Z \in \mathcal{A}^\perp(L_d) \end{cases} \quad (3.7)$$

qui est un système non-linéaire.

La direction de Newton (D_X, D_Z) est donc solution de ce système linéarisé.

Dans un premier temps, linéarisons l'équation $V(1)^2 = \gamma\mu I$.

$$\gamma\mu I = V(1)^2 \approx V(0)^2 + \frac{dV(t)^2}{dt} \Big|_{t=0.1}$$

Calculons $\frac{dV(t)^2}{dt} \Big|_{t=0}$.

Puisque $\frac{dV(t)^2}{dt} = 2\mathcal{P}_H(V(t)V^{(1)}(t))$, calculons dans un premier temps $V^{(1)}(0)$.

Nous savons que $V(t) := G(t)^H Z(t) G(t) = G(t)^{-1} X(t) G(t)^{-H}$, et donc nous avons

$$\begin{aligned} V(t) &= \frac{1}{2} G(t)^H Z(t) G(t) + \frac{1}{2} G(t)^{-1} X(t) G(t)^{-H} \\ V^{(1)}(t) &= \frac{dV(t)}{dt} = \frac{1}{2} \{ G^{(1)}(t)^H Z(t) G(t) + G(t)^H Z^{(1)}(t) G(t) + G(t)^H Z(t) G^{(1)}(t) \\ &\quad + (G(t)^{-1})^{(1)} X(t) G(t)^{-H} + G(t)^{-1} X^{(1)}(t) G(t)^{-H} + G(t)^{-1} X(t) (G(t)^{-H})^{(1)} \} \end{aligned}$$

Or, nous avons

$$\begin{aligned} &\frac{1}{2} \{ G^{(1)}(t)^H Z(t) G(t) + (G(t)^{-1})^{(1)} X(t) G(t)^{-H} + G(t)^H Z^{(1)}(t) G(t) + (G(t)^{-1}) X(t) (G(t)^{-H})^{(1)} \} \\ &= \mathcal{P}_H(G^{(1)}(t)^H Z(t) G(t) + (G(t)^{-1})^{(1)} X(t) G(t)^{-H}) \end{aligned}$$

car $X(t), Z(t) \in \mathcal{H}$ et voir annexe. Il suit que

$$\begin{aligned} V^{(1)}(t) &= \frac{1}{2}G(t)^H Z^{(1)}(t)G(t) + \frac{1}{2}G(t)^{-1}X^{(1)}(t)G(t)^{-H} \\ &\quad + \mathcal{P}_{\mathcal{H}}(G^{(1)}(t)^H Z(t)G(t) + (G(t)^{-1})^{(1)}X(t)G(t)^{-H}). \end{aligned}$$

Or, nous savons que

$$\frac{dG(t)G(t)^{-1}}{dt} = \frac{dI}{dt} = 0 = G(t)^{(1)}G(t)^{-1} + G(t)\frac{dG(t)^{-1}}{dt}.$$

Il suit que

$$\frac{dG(t)^{-1}}{dt} = -G(t)^{-1}G(t)^{(1)}G(t)^{-1} \quad (3.8)$$

et, nous écrivons successivement

$$\begin{aligned} V^{(1)}(t) &= \frac{1}{2}G(t)^H Z^{(1)}(t)G(t) + \frac{1}{2}G(t)^{-1}X^{(1)}(t)G(t)^{-H} \\ &\quad + \mathcal{P}_{\mathcal{H}}(G^{(1)}(t)^H Z(t)G(t) - G(t)^{-1}G(t)^{(1)}G(t)^{-1}X(t)G(t)^{-H}) \\ &= \frac{1}{2}G(t)^H Z^{(1)}(t)G(t) + \frac{1}{2}G(t)^{-1}X^{(1)}(t)G(t)^{-H} \\ &\quad + \mathcal{P}_{\mathcal{H}}(G^{(1)}(t)^H G(t)^{-H}G(t)^H Z(t)G(t) - G(t)^{-1}G(t)^{(1)}G(t)^{-1}X(t)G(t)^{-H}) \\ &= \frac{1}{2}G(t)^H Z^{(1)}(t)G(t) + \frac{1}{2}(G(t)^{-1})X^{(1)}(t)G(t)^{-H} \\ &\quad + \mathcal{P}_{\mathcal{H}}((G^{(1)}(t)^H G(t)^{-H} - G(t)^{-1}G(t)^{(1)})V(t)). \end{aligned}$$

Utilisant le fait que $G(0)=I$, nous obtenons :

$$V^{(1)}(0) = \frac{1}{2}(Z^{(1)}(0) + X^{(1)}(0)) + \mathcal{P}_{\mathcal{H}}((G^{(1)}(0)^H - G^{(1)}(0))V). \quad (3.9)$$

Or, comme nous avons posé $G(t) = D(X(t), Z(t))^{\frac{1}{2}}$, nous avons $G(t) \in \mathcal{H}$ et ceci pour tout t . Dès lors, nous obtenons $G^{(1)}(0) \in \mathcal{H}$ et nous simplifions l'expression de $V^{(1)}(0)$ qui devient

$$V^{(1)}(0) = \frac{1}{2}(Z^{(1)}(0) + X^{(1)}(0)). \quad (3.10)$$

Par conséquent, nous obtenons

$$\begin{aligned} \frac{dV(t)^2}{dt}|_{t=0} &= 2\mathcal{P}_{\mathcal{H}}(V(0)V^{(1)}(0)) \\ &= 2\mathcal{P}_{\mathcal{H}}(V\frac{1}{2}(Z^{(1)}(0) + X^{(1)}(0))) \\ &= \mathcal{P}_{\mathcal{H}}(V(D_X + D_Z)). \end{aligned}$$

Finalement, nous avons

$$V(1)^2 \approx V(0)^2 + \mathcal{P}_{\mathcal{H}}(V(D_X + D_Z))$$

c'est-à-dire

$$\mathcal{P}_{\mathcal{H}}(V(D_X + D_Z)) \approx \gamma\mu I - V^2$$

Le système linéarisé donnant la direction de Newton (D_X, D_Z) est donc

$$\begin{cases} \mathcal{P}_{\mathcal{H}}(V(D_X + D_Z)) = \gamma\mu I - V^2 \\ D_X \in \mathcal{A}(L_d) \\ D_Z \in \mathcal{A}^\perp(L_d) \end{cases}$$

Pour résoudre la première équation, nous nous aidons du résultat suivant :

Lemme 3.2 *Soient V une matrice diagonale à éléments positifs et Y une matrice hermitienne du même ordre*

Alors

$$\mathcal{P}_{\mathcal{H}}(VX) = Y \iff X_{ij} = \frac{2Y_{ij}}{V_{ii} + V_{jj}} \quad \forall i, j.$$

En ce qui nous concerne, nous avons V une matrice diagonale à éléments positifs (voir 3.1.2). Nous voulons $D_X + D_Z$ hermitienne. Par conséquent, la solution unique de $\mathcal{P}_{\mathcal{H}}(V(D_X + D_Z)) = \gamma\mu I - V^2$ est $D_X + D_Z = \gamma\mu V^{-1} - V$.

En effet, le lemme 3.2 nous dit que $(D_X + D_Z)_{ij} = \frac{2(\gamma\mu\delta_{ij} - V_{ij}^2)}{V_{ii} + V_{jj}}$.

Pour le cas où

$$\begin{aligned} - i=j : (D_X + D_Z)_{ii} &= \frac{2(\gamma\mu - V_{ii}^2)}{2V_{ii}} = \gamma\mu V_{ii}^{-1} - V_{ii} \\ - i \neq j : (D_X + D_Z)_{ij} &= \frac{-2V_{ij}^2}{V_{ii} + V_{jj}} = 0 \text{ puisque } V \text{ est diagonale.} \end{aligned}$$

En conclusion, la direction de Newton (D_X, D_Z) au point (\bar{X}, \bar{Z}) dans l'espace transformé satisfait

$$D_X + D_Z = \gamma\mu V^{-1} - V \tag{3.11}$$

ainsi que $D_X \in \mathcal{A}(L_d)$ et $D_Z \in \mathcal{A}^\perp(L_d)$.

Avant de se ramener à l'espace de départ, tirons quelques résultats nécessaires pour la suite.

Tout d'abord, $D_X \in \mathcal{A}(L_d)$ et $D_Z \in \mathcal{A}^\perp(L_d)$. Nous obtenons donc

$$\|D_X\|_F^2 + \|D_Z\|_F^2 = \|D_X + D_Z\|_F^2. \tag{3.12}$$

Ensuite, puisque $V^2 = \mu I$, nous obtenons

$$\begin{aligned} \operatorname{Re} \operatorname{tr}(\mu I - V^2) &= 0 \\ V \bullet (\mu V^{-1} - V) &= 0 \\ V &\perp (\mu V^{-1} - V) \end{aligned}$$

et nous avons successivement

$$\begin{aligned} \|D_X + D_Z\|_F^2 &= \|\gamma(\mu V^{-1} - V) - (1 - \gamma)V\|_F^2 \\ &= \gamma^2 \|\mu V^{-1} - V\|_F^2 + (1 - \gamma)^2 \|V\|_F^2 \\ &= \gamma^2 \|\mu V^{-1} - V\|_F^2 + (1 - \gamma)^2 \bar{n}\mu \end{aligned} \tag{3.13}$$

De plus, le nouveau saut de dualité devient

$$\begin{aligned}
X(t) \bullet Z(t) &= \operatorname{Re} \operatorname{tr}(V + tD_X)(V + tD_Z) \\
&= \operatorname{Re} \operatorname{tr}(V^2 + tV(D_X + D_Z) + t^2(D_X D_Z)) \\
&= \operatorname{Re} \operatorname{tr} V^2 + t \operatorname{Re} \operatorname{tr}(V(D_X + D_Z)) \quad \text{car } D_X \perp D_Z \\
&= \|V\|_F^2 + tV \bullet (\gamma\mu V^{-1} - V) \\
&= \|V\|_F^2(1 - t) + tV \bullet \gamma\mu V^{-1} \\
&= \|V\|_F^2(1 - t) + t\gamma \operatorname{Re} \operatorname{tr} V^2 \quad \text{car } \mu I = V^2,
\end{aligned}$$

ce qui mène à

$$X(t) \bullet Z(t) = (1 - t + \gamma t)\|V\|_F^2. \quad (3.14)$$

Comme $X(t) \bullet Z(t) = \|V(t)\|_F^2$, nous obtenons $\|V(t)\|_F^2 = (1 - t + \gamma t)\|V\|_F^2$. Ainsi, tout comme $\mu = \frac{\|V\|_F^2}{\bar{n}}$, nous pouvons définir

$$\mu(t) := \frac{\|V(t)\|_F^2}{\bar{n}} = (1 - t + \gamma t)\mu. \quad (3.15)$$

3.2.2 Calcul de la direction de Newton en un point primal-dual (X,Z) de l'espace initial

Soient

$X(0) \in (B + \mathcal{A}) \cap \mathcal{H}_{\bar{n}}^{++}$ un point intérieur de l'espace primal initial,
 $Z(0) \in (C + \mathcal{A}^\perp) \cap \mathcal{H}_{\bar{n}}^{++}$ un point intérieur de l'espace dual initial,
 $(X(0), Z(0))$ le point correspondant de l'espace transformé.

Soient (D_X, D_Z) la direction de Newton au point primal-dual $(X(0), Z(0))$. Cette direction peut être exprimée dans l'espace de départ comme suivant :

$$\begin{aligned}
\Delta X &= L_d D_X L_d^H \in \mathcal{A} \quad \text{car } D_X = L_d^{-1} \Delta X L_d^{-H} \\
\Delta Z &= L_d^{-H} D_Z L_d^{-1} \in \mathcal{A}^\perp \quad \text{car } D_Z = L_d^H \Delta Z L_d.
\end{aligned}$$

Les équations de Newton permettant de calculer les directions de Newton (D_X, D_Z) dans l'espace transformé peuvent également être exprimées dans l'espace de départ comme suivant :

$$\begin{cases} \Delta X + D(X, Z) \Delta Z D(X, Z) = \gamma\mu Z^{-1} - X \\ \Delta X \in \mathcal{A} \text{ et } \Delta Z \in \mathcal{A}^\perp. \end{cases}$$

En effet, nous savons d'une part que

$$\begin{aligned}
D_X + D_Z &= \gamma\mu V^{-1} - V \\
L_d(D_X + D_Z)L_d^H &= L_d(\gamma\mu V^{-1} - V)L_d^H \\
\Delta X + L_d D_Z L_d^H &= \gamma\mu L_d V^{-1} L_d^H - L_d V L_d^H \quad \text{par définition de } \Delta X
\end{aligned}$$

Or, nous écrivons

$$\begin{aligned}
L_d D_Z L_d^H &= L_d L_d^H L_d^{-H} D_Z L_d^{-1} L_d L_d^H \\
&= L_d L_d^H \Delta Z L_d L_d^H \quad \text{par définition de } \Delta Z \\
&= D(X, Z) \Delta Z D(X, Z) \quad \text{voir 3.1.2 (lemme 3.1)}
\end{aligned}$$

et nous avons $V = L_d^{-1} X L_d^{-H} = L_d^H Z L_d$. Par conséquent, nous obtenons que $\Delta X + D(X, Z) \Delta Z D(X, Z) = \gamma \mu Z^{-1} - X$.

D'autre part, nous avons $\Delta X \in \mathcal{A}$ et $\Delta Z \in \mathcal{A}^\perp$, ce qui nous mène au résultat.

3.3 Voisinages de la trajectoire centrale

Nous avons mentionné au début du point 3.2.1. que, dans le calcul de la direction de Newton, nous ne tenions pas compte du caractère défini positif devant être impérativement vérifié par nos itérés. Nous nous en préoccupons dans cette section-ci.

Un avantage de la matrice de transformation L_d relevé dans le point 3.1.3 réside dans le caractère défini positif des points se trouvant dans un voisinage de la trajectoire centrale. C'est pourquoi, dans nos futurs algorithmes de suivi de chemin, nous choisirons γ le paramètre de centrage et t la longueur de pas de telle manière que, partant d'un point $(X(0), Z(0))$ proche de la trajectoire centrale, l'itéré suivant $(X(t), Z(t))$ soit également proche de cette trajectoire centrale.

Afin de quantifier cette proximité, nous introduisons une mesure de la distance à la trajectoire centrale :

$$\begin{aligned}
\delta : \mathcal{H}_{\bar{n}} &\longrightarrow \mathbb{R}^+ \\
Y &\longrightarrow \delta(Y) := \left\| I - \frac{\bar{n}}{I \bullet Y^2} Y^2 \right\|_F \\
&= \left| \frac{\bar{n}}{I \bullet Y^2} \right| \left\| \frac{I \bullet Y^2}{\bar{n}} I - Y^2 \right\|_F.
\end{aligned}$$

En particulier, nous avons

$$\delta(V(t)) = \left\| I - \frac{\bar{n}}{I \bullet V(t)^2} V(t)^2 \right\|_F \quad (3.16)$$

Or, nous savons que $I \bullet V(t)^2 = \text{Re tr} V(t)^2 = \|V(t)\|_F^2 = \mu(t) \bar{n}$. Il suit que $\delta(V(t)) = \left\| I - \frac{1}{\mu(t)} V(t)^2 \right\|_F$.

Nous voyons donc que si la distance $\delta(Y)$ est nulle, alors Y se trouve sur la trajectoire centrale, c'est-à-dire que Y est un multiple de l'identité.

Nous basant sur cette mesure, nous définissons des voisinages de la trajectoire centrale auxquels nous imposerons à nos itérés d'appartenir. De cette manière, nos itérés seront

tous proches de la trajectoire centrale et définis positifs.

Le premier semble naturel et se définit comme ceci :

$$\mathcal{N}_2(\beta) := \{Y \in \mathcal{H}_{\bar{n}}^{++} | \delta(Y) \leq \beta\} \quad \text{où } \beta \text{ est une constante } \in]0, 1[. \quad (3.17)$$

Une propriété importante de ce voisinage est :

$$\text{Si } V(t) \in \mathcal{N}_2(\beta), \text{ alors } (1 + \beta)\mu(t)I \succeq V(t)^2 \succeq (1 - \beta)\mu(t)I. \quad (3.18)$$

En effet, nous écrivons successivement

$$\begin{aligned} V(t) \in \mathcal{N}_2(\beta) &\iff \delta(V(t)) \leq \beta \\ &\iff \left\| I - \frac{1}{\mu(t)} V(t)^2 \right\|_F \leq \beta \\ &\iff \left(\sum_{i=1}^{\bar{n}} \left(1 - \frac{\lambda_i(t)}{\mu(t)} \right)^2 \right)^{\frac{1}{2}} \leq \beta \\ &\quad \text{où } \lambda_i(t) \text{ } i = 1 \dots \bar{n} \text{ est une valeur propre de } V(t)^2 \\ &\implies \left| 1 - \frac{\lambda_i(t)}{\mu(t)} \right| \leq \beta \text{ } i = 1 \dots \bar{n} \\ &\iff -\beta \leq 1 - \frac{\lambda_i(t)}{\mu(t)} \leq \beta \text{ } i = 1 \dots \bar{n} \\ &\iff 1 - \beta \leq \frac{\lambda_i(t)}{\mu(t)} \leq 1 + \beta \text{ } i = 1 \dots \bar{n} \end{aligned}$$

Or, si toutes les valeurs propres d'une matrice sont positives, alors cette matrice est semi-définie positive. Ainsi, la thèse est obtenue.

Le second voisinage de la trajectoire centrale est défini comme suit :

$$\mathcal{N}_{\infty}^{-} := \{Y \in \mathcal{H}_{\bar{n}}^{++} | Y^2 \succeq (1 - \beta) \frac{I \bullet Y^2}{\bar{n}} I\} \quad \text{où } \beta \text{ est une constante } \in]0, 1[. \quad (3.19)$$

Ce voisinage est plus large que le premier. En termes mathématiques, nous écrivons $\mathcal{N}_2(\beta) \subseteq \mathcal{N}_{\infty}^{-}(\beta)$.

En effet, nous avons

$$\begin{aligned} V(t) \in \mathcal{N}_2(\beta) &\implies V(t)^2 \succeq (1 - \beta)\mu(t)I \quad (\text{par (3.18)}) \\ \text{Or, } \mu(t) &= \frac{\|V(t)^2\|_F}{\bar{n}} \\ &\implies V(t)^2 \succeq (1 - \beta) \frac{I \bullet V(t)^2}{\bar{n}} I \\ &\implies V(t) \in \mathcal{N}_{\infty}^{-}(\beta). \end{aligned}$$

Ceci explique pourquoi nous nommons $\mathcal{N}_{\infty}^{-}(\beta)$: voisinage large.

3.4 Lemmes techniques

Dans cette section, nous citons les lemmes techniques indispensables à l'obtention des résultats importants de ce chapitre. Ceux-ci ne sont pas démontrés, comme certains autres lemmes de ce mémoire. Ceci résulte d'un choix de ma part, privilégiant une vue d'ensemble des méthodes de suivi de chemin en programmation semi-définie.

Lemme 3.3 Soit $\gamma \in]0, 1]$ et $\beta \in]0, 1[$.

Si $V \in \mathcal{N}_2(\beta)$ et $\frac{\beta^2}{1-\beta} + \bar{n}(\frac{1-\gamma}{\gamma})^2 \leq 1$,

alors $t^* > 1$ et $V(1) \succ 0$ où $t^* = \max\{t | V + tD_X \succeq 0 \text{ et } V + tD_Z \succeq 0\}$.

Lemme 3.4 Soit $\delta(V) < 1$. Pour $0 \leq t < t^*$, on a

$$(1 - t + \gamma t)\delta(V(t)) \leq (1 - t)\delta(V) + \frac{t^2}{2}(\frac{\gamma^2\delta(V)^2}{1-\delta(V)} + \bar{n}(1 - \gamma)^2).$$

En supposant, $V \in \mathcal{N}_2(\beta)$ et $\gamma \in]0, 1]$, cette relation devient : $(1 - t + \gamma t)\delta(V(t)) \leq (1 - t)\beta + \frac{1}{2}(\gamma t)^2(\frac{\beta^2}{1-\beta} + \bar{n}(\frac{1-\gamma}{\gamma})^2)$.

Lemme 3.5 Si $\gamma = 1$ et $\delta(V) \leq \frac{1}{2}$,

alors $\delta(V(1)) \leq \delta(V)^2$.

Lemme 3.6 Soit $0 < \gamma < 1$ et $V \in \mathcal{N}_\infty^-(\beta)$ où $\beta \in]0, 1[$.

Alors, on a $V(t) \in \mathcal{N}_\infty^-(\beta) \forall t \geq 0$ tel que $0 \leq \bar{n}t \leq \frac{2\beta\gamma}{\frac{\beta\gamma^2}{1-\beta} + (1-\gamma^2)}$.

Lemme 3.7 Soit $\delta(V) < 1$. Pour $0 \leq t < t^*$, on a

$$\mu(t)\delta(V(t)) \leq (1 - t)\mu\delta(V) + t^2\|\mathcal{P}_\mathcal{H}(D_X^p D_Z^p)\|_F.$$

La preuve de ces lemmes se trouve respectivement aux pages 65, 65, 66 et 66 de [14].

3.5 Algorithmes de suivi de chemin

A présent, puisque l'espace transformé nous a permis de calculer les équations nous donnant la direction de Newton (D_X, D_Z) en un point (X, Z) dans l'espace initial, nous nous restreignons à l'espace de départ.

Nous savons que, partant d'un itéré appartenant à un voisinage de la trajectoire centrale, nous allons par nos algorithmes de suivi de chemin engendrer des itérés proches de la trajectoire dont le saut de dualité diminue. De cette manière, les itérés convergeront vers une solution optimale. Ceux-ci seront calculés grâce à la direction de Newton (voir 3.2.2) et grâce à une longueur de pas t à déterminer. Ces considérations vont nous permettre dans cette section de construire dans un premier temps un algorithme général de suivi de chemin dont les paramètres encore à fixer sont $t^{(k)}$ la longueur de pas de la $k^{ième}$ itération ($k=0,1,\dots$), $\gamma^{(k)}$ le paramètre de centrage de la $k^{ième}$ itération ($k=0,1,\dots$), et β le paramètre du voisinage sélectionné. Ces différents choix nous mèneront ensuite à quatre algorithmes différents, tous étant des généralisations d'algorithmes de la programmation linéaire. Nous étudierons le comportement de leur saut de dualité ainsi que leur complexité algorithmique.

3.5.1 Algorithme général de suivi de chemin

Cet algorithme calcule une solution ϵ -optimale.

Etant donnés $\epsilon > 0$ et $(X^{(0)}, Z^{(0)})$ point initial tel que $V(0) \in \mathcal{N}(\beta)$,

PAS 0 poser $k := 0$

PAS 1 si $X^{(k)} \bullet Z^{(k)} < \epsilon$, alors l'algorithme se termine puisque le saut de dualité est assez petit.

PAS 2 choisir $\gamma^{(k)} \in [0, 1]$ paramètre de centrage et résoudre le système linéaire suivant, appelé équations de Newton, afin de calculer la direction de Newton $(\Delta X, \Delta Z)$:

$$\begin{cases} \Delta X^{(k)} + D(X^{(k)}, Z^{(k)}) \Delta Z^{(k)} D(X^{(k)}, Z^{(k)}) = \gamma^{(k)} \mu(Z^{(k)})^{-1} - X^{(k)} \\ \Delta X^{(k)} \in \mathcal{A} \text{ et } \Delta Z^{(k)} \in \mathcal{A}^\perp \end{cases} \quad (3.20)$$

où $\mu^{(k)} = \frac{X^{(k)} \bullet Z^{(k)}}{\bar{n}}$

PAS 3 choisir $t^{(k)}$

poser $X^{(k+1)} := X^{(k)} + t^{(k)} \Delta X^{(k)}$ et $Z^{(k+1)} := Z^{(k)} + t^{(k)} \Delta Z^{(k)}$

PAS 4 poser $k := k+1$ et retourner au pas 1

De la manière dont nous avons construit cet algorithme, nous devons choisir impérativement nos paramètres de telle façon que $V^{(k)} \in \mathcal{N}(\beta) \quad \forall k$.

3.5.2 Algorithme à petits pas

Tout comme pour l'algorithme à petits pas pour la programmation linéaire, nous choisissons $\mathcal{N}(\beta) = \mathcal{N}_2(\frac{1}{2})$, ainsi que $t^{(k)} = 1$ pour $k=0,1,\dots$ et $\gamma^{(k)} = \gamma$ pour $k=0,1,\dots$ avec, dans notre cas, $\gamma = \frac{1}{1 + \frac{1}{\sqrt{2\bar{n}}}}$.

Montrons tout d'abord que ce choix de paramètres impose à $V^{(k)}$ d'appartenir à $\mathcal{N}_2(\beta)$ pour $k=0,1,\dots$. Démontrons ceci par récurrence.

Nous savons grâce à l'algorithme général que $V^{(0)} \in \mathcal{N}_2(\beta)$.

Soit $k \in \mathbb{N}$. Montrons que $V^{(k)} \in \mathcal{N}_2(\beta)$ entraîne que $V^{(k+1)} \in \mathcal{N}_2(\beta)$.

En fait, notre choix de paramètres nous permet de vérifier les hypothèses du lemme 3.3 : $\gamma \in]0, 1]$, $\beta \in]0, 1]$, $V^{(k)} \in \mathcal{N}_2(\beta)$, et

$$\begin{aligned} \frac{\beta^2}{1-\beta} + \bar{n} \left(\frac{1-\gamma}{\gamma} \right)^2 &= \frac{\frac{1}{4}}{\frac{1}{2}} + \bar{n} \left(\frac{1 + \frac{1}{\sqrt{2\bar{n}}} - 1}{1 + \frac{1}{\sqrt{2\bar{n}}}} \right)^2 \left(1 + \frac{1}{\sqrt{2\bar{n}}} \right)^2 \\ &= \frac{1}{2} + \frac{\bar{n}}{2\bar{n}} = 1. \end{aligned}$$

Nous pouvons donc conclure que le pas de longueur maximale t^* est strictement supérieur à 1 et que $V(1) \succ 0$. De plus, ayant choisi $V^{(k)} \in \mathcal{N}_2(\beta)$, nous satisfaisons l'hypothèse du lemme 3.4 ($\delta(V^{(k)}) < 1$) et obtenons donc $\forall 0 \leq t \leq t^*$:

$$(1-t+\gamma t)\delta(V^{(k+1)}(t)) \leq (1-t)\delta(V^{(k)}) + \frac{t^2}{2} \left(\frac{\gamma^2 \delta(V^{(k)})^2}{1-\delta(V^{(k)})} + \bar{n}(1-\gamma)^2 \right).$$

Comme $t^* > 1$ et $t^{(k)}=1 \forall k$, nous remplaçons t par $t^{(k)}$ dans cette expression, cela nous donne

$$(1 - 1 + \gamma 1) \delta(V^{(k+1)}(1)) \leq (1 - 1) \beta + \frac{1}{2} (\gamma^2 1^2) \left(\frac{\beta^2}{1 - \beta} + \bar{n} \left(\frac{1 - \gamma}{\gamma} \right)^2 \right)$$

expression équivalente à

$$\gamma \delta(V^{(k+1)}) \leq \frac{1}{2} \gamma^2 \cdot 1$$

Or, nous savons que $\gamma < 1$. Il suit que $\delta(V^{(k+1)}) < \frac{1}{2} = \beta$.

Dans un second temps, analysons le comportement du saut de dualité dans cet algorithme à petits pas. Nous avons successivement,

$$\begin{aligned} X^{(k+1)} \bullet Z^{(k+1)} &= \bar{n} \mu^{(k+1)} \quad \text{par définition dans l'algorithme général} \\ &= \bar{n} \gamma \mu^{(k)} \quad \text{par (3.15) et car } t^{(k)} = 1 \forall k \\ &= \gamma X^{(k)} \bullet Z^{(k)}. \end{aligned}$$

Or, $\gamma \in]0, 1[$ car $\gamma = \frac{1}{1 + \frac{1}{\sqrt{2\bar{n}}}}$ où \bar{n} est la dimension des matrices traitées. Donc, $X^{(k+1)} \bullet Z^{(k+1)}$ converge vers 0 lorsque k tend vers $+\infty$.

Dans un troisième temps, intéressons-nous à la complexité de cet algorithme.

Théorème 3.1 *L'algorithme à petits pas calcule une solution ϵ -optimale en $O(\sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$ itérations.*

preuve :

Nous savons que

$$\begin{aligned} X^{(k+1)} \bullet Z^{(k+1)} &= \gamma X^{(k)} \bullet Z^{(k)} \\ \log(X^{(k+1)} \bullet Z^{(k+1)}) &= \log \gamma + \log(X^{(k)} \bullet Z^{(k)}) \\ &= \log \gamma + \log \gamma + \log(X^{(k-1)} \bullet Z^{(k-1)}) \\ &= (k+1) \log \gamma + \log(X^{(0)} \bullet Z^{(0)}) \\ &= -(k+1) \log \left(1 + \frac{1}{\sqrt{2\bar{n}}} \right) + \log(X^{(0)} \bullet Z^{(0)}) \\ &\leq -\frac{(k+1)}{2\sqrt{2\bar{n}}} + \log(X^{(0)} \bullet Z^{(0)}) \end{aligned}$$

car

$$\begin{aligned} -\log \left(1 + \frac{1}{\sqrt{2\bar{n}}} \right) &\leq -\frac{1}{2\sqrt{2\bar{n}}} \\ \iff \log(1+x) &\geq \frac{x}{2} \quad \text{où } x = \frac{1}{\sqrt{2\bar{n}}} \text{ et } 0 < x < 1 \\ \iff 1+x &\geq e^{\frac{x}{2}} \quad \text{où } 0 < x < 1 \end{aligned}$$

qui est vérifié. Par conséquent, nous avons

$$\log(X^{(k+1)} \bullet Z^{(k+1)}) \leq -\frac{(k+1)}{2\sqrt{2\bar{n}}} + \log(X^{(0)} \bullet Z^{(0)}).$$

Or, $(X^{(k+1)}, Z^{(k+1)})$ est une solution ϵ -optimale

$$\begin{aligned} &\iff X^{(k+1)} \bullet Z^{(k+1)} \leq \epsilon \\ &\iff X^{(k+1)} \bullet Z^{(k+1)} \leq X^{(0)} \bullet Z^{(0)} e^{-\frac{k+1}{2\sqrt{2n}}} = \epsilon. \end{aligned}$$

De par cette dernière équation, nous trouvons

$$k + 1 = 2\sqrt{2n} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon}.$$

Ceci nous mène à dire que

$$k + 1 = O(\sqrt{n} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon}).$$

cqfd

Nous devons cependant constater que cet algorithme est en fait peu pratique vu son taux de convergence du saut de dualité fixé. En effet, $\gamma^{(k)} = \gamma \quad \forall k$ et $X^{(k+1)} \bullet Z^{(k+1)} = \gamma X^{(k)} \bullet Z^{(k)}$. De plus, la longueur de pas est également fixée : $t^{(k)} = 1$ pour tout k , ce qui freine l'algorithme.

Nous allons donc essayer de pallier à ces deux inconvénients en proposant l'algorithme suivant.

3.5.3 Algorithme prédicteur-correcteur

Tout comme pour l'algorithme prédicteur-correcteur de la programmation linéaire, nous choisissons $\mathcal{N}(\beta) = \mathcal{N}_2(\frac{1}{2})$, ainsi que, d'une part

$$\gamma^{(k)} = \begin{cases} 1 & \text{si } k \text{ est pair} \\ 0 & \text{si } k \text{ est impair} \end{cases}$$

et d'autre part,

$$t^{(k)} = \begin{cases} 1 & \text{si } k \text{ est pair} \\ \max\{\bar{t} \mid V^{(k)}(t) \in \mathcal{N}_2(\beta) \text{ pour } 0 \leq t \leq \bar{t}\} & \text{si } k \text{ est impair.} \end{cases}$$

Nous obtenons de ce fait l'algorithme suivant :

Etant donnés $\epsilon > 0$ et $(X^{(0)}, Z^{(0)})$ point initial où $V(0) \in \mathcal{N}_2(\frac{1}{2})$,

PAS 0 poser $k := 0$

PAS 1 si $X^{(k)} \bullet Z^{(k)} < \epsilon$, alors l'algorithme se termine puisque le saut de dualité est assez petit.

PAS 2 - si k est pair : PAS CORRECTEUR :

$\gamma^{(k)} = 1$ paramètre de centrage;

calculer la direction de Newton $(\Delta X, \Delta Z)$ en résolvant le système linéaire des équations de Newton (3.20);

$t^{(k)} = 1$ la longueur de pas.

- si k est impair : PAS PREDICTEUR :

$\gamma^{(k)} = 0$ paramètre de centrage;

calculer la direction de Newton $(\Delta X, \Delta Z)$ en résolvant le système linéaire des équations de Newton (3.20);

$t^{(k)} = \max\{\bar{t} | V^{(k)}(t) \in \mathcal{N}_2(\beta) \text{ pour } 0 \leq t \leq \bar{t}\}$ la longueur de pas.

PAS 3 poser $X^{(k+1)} := X^{(k)} + t^{(k)} \Delta X^{(k)}$ et $Z^{(k+1)} := Z^{(k)} + t^{(k)} \Delta Z^{(k)}$

PAS 4 poser $k := k+1$ et retourner au pas 1

Explicitons quelque peu ce que nous faisons lorsque nous effectuons un pas correcteur ou un pas prédicteur, c'est-à-dire ce qu'il advient du saut de dualité et de la contrainte sur les itérés qui est de rester dans le voisinage $\mathcal{N}_2(\frac{1}{2})$.

Occupons-nous tout d'abord du pas correcteur (k est pair). Celui-ci conserve le saut de dualité mais améliore la centralité de l'itéré.

En effet, soit $(X^{(k)}, Z^{(k)})$ l'itéré au début du pas correcteur et soit $(X^{(k+1)}, Z^{(k+1)})$ l'itéré à la fin du pas correcteur.

- Du point de vue du saut de dualité, nous avons successivement les égalités suivantes :

$$\begin{aligned} X^{(k+1)} \bullet Z^{(k+1)} &= \bar{n} \mu^{(k+1)} \text{ par définition dans l'algorithme} \\ &= \bar{n} \mu^{(k)} \text{ par (3.15) et le fait que } t^{(k)} = \gamma^{(k)} = 1 \\ &= X^{(k)} \bullet Z^{(k)}. \end{aligned}$$

Ceci indique la conservation du saut de dualité lors d'un pas correcteur.

- Du point de vue de la centralité, nous avons que si $V^{(k)} \in \mathcal{N}_2(\frac{1}{2})$ alors $V^{(k+1)} \in \mathcal{N}_2(\frac{1}{4})$ car, il suit du lemme 3.5 que $\delta(V^{(k+1)}) \leq \delta(V^{(k)})^2 \leq \frac{1}{4}$.

Ceci indique que l'itéré se rapproche de la trajectoire centrale lors d'un pas correcteur.

Par contre, en nous intéressant ensuite au pas prédicteur (k est impair), nous voyons que celui-ci réduit le saut de dualité mais n'améliore pas la centralité de l'itéré.

En effet, soit $(X^{(k)}, Z^{(k)})$ l'itéré au début du pas prédicteur et soit $(X^{(k+1)}, Z^{(k+1)})$ l'itéré à la fin du pas prédicteur.

- Du point de vue de la centralité, nous avons que si $V^{(k)} \in \mathcal{N}_2(\frac{1}{4})$ alors $V^{(k+1)} \in \mathcal{N}_2(\frac{1}{2})$ car, par le lemme 3.4, nous avons successivement avec $\gamma=0$

$$\begin{aligned} (1-t)\delta(V^{(k+1)}(t)) &\leq (1-t)\delta(V^{(k)}) + \frac{t^2}{2}\bar{n} \\ \iff \delta(V^{(k+1)}(t)) &\leq \delta(V^{(k)}) + \frac{t^2}{2}\frac{\bar{n}}{(1-t)} \\ \iff \delta(V^{(k+1)}(t)) &\leq \frac{1}{4} + \frac{t^2}{(1-t)}\frac{\bar{n}}{2}. \end{aligned}$$

Ceci veut dire que si $0 \leq t \leq \frac{2}{1+\sqrt{1+8\bar{n}}}$ alors $\delta(V^{(k+1)}(t)) \leq \frac{1}{2}$, puisque, pour $t = \frac{2}{1+\sqrt{1+8\bar{n}}}$, nous avons $\delta(V^{(k+1)}(t)) \leq \frac{1}{2}$ et car $\frac{1}{4} + \frac{t^2}{(1-t)^2} \frac{\bar{n}}{2}$ est une fonction croissante en t sur $[0, \frac{2}{1+\sqrt{1+8\bar{n}}}]$ puisque la dérivée de cette fonction, $\frac{\bar{n}t}{(1-t)^2} + \frac{\bar{n}t^2}{2(1-t)^3}$ est positive sur cet intervalle. C'est pourquoi, par la définition de $t^{(k)}$ pour les itérations prédictes, celui-ci sera toujours plus grand ou égal à $\frac{2}{1+\sqrt{1+8\bar{n}}}$.

– Du point de vue du saut de dualité, nous avons successivement les égalités suivantes :

$$\begin{aligned} X^{(k+1)} \bullet Z^{(k+1)} &= \bar{n}\mu^{(k+1)} \quad \text{par définition dans l'algorithme} \\ &= \bar{n}(1 - t^{(k)})\mu^{(k)} \quad \text{par (3.15) et le fait que } \gamma^{(k)} = 0 \\ &= (1 - t^{(k)})X^{(k)} \bullet Z^{(k)}. \end{aligned}$$

Le saut de dualité diminue ainsi strictement à chaque pas prédictes car $t^{(k)} \geq \frac{2}{1+\sqrt{1+8\bar{n}}} > 0$.

Terminons en nous intéressant à la complexité de cet algorithme.

Théorème 3.2 *L'algorithme prédictes-correctes calcule une solution ϵ -optimale en $O(\sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$ itérations. Ainsi, l'algorithme prédictes-correctes possède la même complexité algorithmique que l'algorithme à petits pas.*

preuve : Nous savons que $X^{(k+1)} \bullet Z^{(k+1)} = X^{(k)} \bullet Z^{(k)}$ pour le pas correctes (k pair) et $X^{(k+1)} \bullet Z^{(k+1)} = (1 - t^{(k)})X^{(k)} \bullet Z^{(k)}$ où $t^{(k)} > 0$ pour le pas prédictes (k impair).

De plus, pour le pas prédictes, $t^{(k)} \geq \frac{2}{1+\sqrt{1+8\bar{n}}} > 0$ entraîne que

$$\begin{aligned} X^{(k+1)} \bullet Z^{(k+1)} &\leq (1 - \frac{2}{1+\sqrt{1+8\bar{n}}})X^{(k)} \bullet Z^{(k)} \\ &= (\frac{1}{1+\frac{2}{\sqrt{1+8\bar{n}}-1}})X^{(k)} \bullet Z^{(k)}. \end{aligned}$$

Or, pour l'algorithme à petits pas, nous avons $X^{(k+1)} \bullet Z^{(k+1)} = (\frac{1}{1+\frac{1}{\sqrt{2\bar{n}}}})X^{(k)} \bullet Z^{(k)}$ et une complexité égale à $O(\sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$ pour obtenir une solution ϵ -optimale.

Par conséquent, si nous ne considérons que le pas prédictes, nous pouvons dire que l'algorithme a une complexité égale à celle de l'algorithme à petits pas, $O(\sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$.

En conclusion, puisque le pas correctes ne modifie pas le saut de dualité, l'algorithme prédictes-correctes a besoin de deux fois le nombre d'itérations de la méthode à petits pas pour obtenir une solution ϵ -optimale. En d'autres mots, la complexité de l'algorithme prédictes-correctes est égale à celle de l'algorithme à petits pas, c'est-à-dire égale à $O(\sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$.

cqfd

Nous remarquons que nous pouvons améliorer l'algorithme prédictes-correctes en combinant le pas prédictes et le pas correctes en une seule itération. Nous envisageons cette possibilité avec l'algorithme au plus grand pas.

3.5.4 Algorithme au plus grand pas

Tout comme l'algorithme prédicteur-correcteur, nous choisissons pour voisinage $\mathcal{N}(\beta) = \mathcal{N}_2(\frac{1}{2})$. Cependant, à chaque itération, nous considérons la direction de Newton $(\Delta X^{(k)}, \Delta Z^{(k)})$ comme une combinaison convexe d'une direction centrante l'itéré $(\Delta X_C^{(k)}, \Delta Z_C^{(k)})$ et d'une direction prédisant une solution optimale, $(\Delta X_P^{(k)}, \Delta Z_P^{(k)})$.

En effet, de manière identique au calcul de la direction "correcteur" dans l'algorithme prédicteur-correcteur (cas où k est pair), nous posons $\gamma^{(k)} = 1$ dans les équations de Newton, ainsi nous obtenons $(\Delta X_C^{(k)}, \Delta Z_C^{(k)})$ en résolvant le système linéaire :

$$\begin{cases} \Delta X_C^{(k)} + D(X^{(k)}, Z^{(k)}) \Delta Z_C^{(k)} D(X^{(k)}, Z^{(k)}) = \mu(Z^{(k)})^{-1} - X^{(k)} \\ \Delta X_C^{(k)} \in \mathcal{A} \text{ et } \Delta Z_C^{(k)} \in \mathcal{A}^\perp \end{cases}$$

En ce qui concerne la direction $(\Delta X_P^{(k)}, \Delta Z_P^{(k)})$, nous procédons de manière identique au calcul de la direction "prédicteur" dans le cas de l'algorithme précédent (cas où k est impair), c'est-à-dire que nous posons $\gamma^{(k)} = 0$ dans les équations de Newton et obtenons $(\Delta X_P^{(k)}, \Delta Z_P^{(k)})$ en résolvant ce système linéaire :

$$\begin{cases} \Delta X_P^{(k)} + D(X^{(k)}, Z^{(k)}) \Delta Z_P^{(k)} D(X^{(k)}, Z^{(k)}) = -X^{(k)} \\ \Delta X_P^{(k)} \in \mathcal{A} \text{ et } \Delta Z_P^{(k)} \in \mathcal{A}^\perp \end{cases}$$

De plus, nous définissons la longueur de pas à la $k^{\text{ième}}$ itération $t^{(k)} = 1 \quad \forall k$ ainsi que le paramètre de centrage à la $k^{\text{ième}}$ itération comme

$$\gamma^{(k)} := \min\{\bar{\gamma} | \delta(V_\gamma^{(k+1)}) \leq \beta \text{ pour } \bar{\gamma} \leq \gamma \leq 1\} \quad \forall k.$$

Ainsi, la direction de Newton à chaque itération est définie comme :

$$\begin{aligned} \Delta X_\gamma^{(k)} &= \gamma^{(k)} \Delta X_C^{(k)} + (1 - \gamma^{(k)}) \Delta X_P^{(k)} \\ \Delta Z_\gamma^{(k)} &= \gamma^{(k)} \Delta Z_C^{(k)} + (1 - \gamma^{(k)}) \Delta Z_P^{(k)} \end{aligned}$$

Nous voyons par ceci que $\gamma^{(k)}$ permet un compromis entre une direction "correcteur" et une "direction prédicteur".

Vérifions que $\Delta(X_\gamma^{(k)}, \Delta Z_\gamma^{(k)})$ respecte les équations de Newton. En effet, nous avons successivement les égalités suivantes :

$$\begin{aligned} & \Delta X_\gamma^{(k)} + D(X^{(k)}, Z^{(k)}) \Delta Z_\gamma^{(k)} D(X^{(k)}, Z^{(k)}) \\ = & \gamma^{(k)} \Delta X_C^{(k)} + (1 - \gamma^{(k)}) \Delta X_P^{(k)} + D(X^{(k)}, Z^{(k)}) (\gamma^{(k)} \Delta Z_C^{(k)} + (1 - \gamma^{(k)}) \Delta Z_P^{(k)}) D(X^{(k)}, Z^{(k)}) \\ = & \gamma^{(k)} (\mu(Z^{(k)})^{-1} - X^{(k)}) + (1 - \gamma^{(k)}) (-X^{(k)}) \\ = & \gamma^{(k)} \mu(Z^{(k)})^{-1} - X^{(k)}. \end{aligned}$$

Tout d'abord, nous constatons que la contrainte sur les itérés d'appartenance de $V^{(k)}$ au voisinage $\mathcal{N}_2(\frac{1}{2})$ à chaque itération est respectée grâce au choix de $\gamma^{(k)}$ effectué.

Ensuite, examinons le comportement du saut de dualité de cet algorithme. Celui-ci se note successivement

$$\begin{aligned} X^{(k+1)} \bullet Z^{(k+1)} &= \bar{n} \mu^{(k+1)} \quad \text{par définition dans l'algorithme} \\ &= \bar{n} \gamma^{(k)} \mu^{(k)} \quad \text{par (3.15) et le fait que } t^{(k)} = 1 \quad \forall k \\ &= \gamma^{(k)} X^{(k)} \bullet Z^{(k)}. \end{aligned}$$

Pour avoir la convergence vers 0 du saut de dualité, nous devons montrer que $\gamma^{(k)} \in]0, 1[$. Par définition, nous avons $\gamma^{(k)} := \min\{\bar{\gamma} | \delta(V_{\bar{\gamma}}^{(k+1)}) \leq \beta \text{ pour } \bar{\gamma} \leq \gamma \leq 1\}$. Donc, en montrant que pour $\bar{\gamma} = \frac{1}{1 + \frac{1}{\sqrt{2n}}}$ nous avons $\delta(V_{\bar{\gamma}}^{(k)}) \leq \frac{1}{2}$ pour $\bar{\gamma} \leq \gamma \leq 1$, nous obtiendrons $\gamma^{(k)} \leq \bar{\gamma} = \frac{1}{1 + \frac{1}{\sqrt{2n}}} < 1$.

Soit γ tel que $\bar{\gamma} \leq \gamma \leq 1$, montrons que $\delta(V_{\bar{\gamma}}^{(k)}) \leq \frac{1}{2}$. Nous devons en fait effectuer la même preuve que celle réalisée dans l'analyse de l'algorithme à petits pas mais non plus en considérant $\gamma = \bar{\gamma}$ mais en prenant γ tel que $\bar{\gamma} \leq \gamma \leq 1$. Si nous regardons de plus près cette démonstration par récurrence, nous remarquons que celle-ci reste valable. Seule la dernière hypothèse du lemme (3.3) doit être vérifiée, c'est-à-dire $\frac{\beta^2}{1-\beta} + \bar{n}(\frac{1-\gamma}{\gamma})^2 \leq 1$. Or, pour $\bar{\gamma}$, $\frac{\beta^2}{1-\beta} + \bar{n}(\frac{1-\bar{\gamma}}{\bar{\gamma}})^2 = 1$. Donc, pour $\bar{\gamma} \leq \gamma$, nous avons $\frac{1-\gamma}{\gamma} \leq \frac{1-\bar{\gamma}}{\bar{\gamma}}$. Par conséquent, pour $\bar{\gamma} \leq \gamma$, nous avons $\frac{\beta^2}{1-\beta} + \bar{n}(\frac{1-\gamma}{\gamma})^2 \leq 1$. Ce développement nous permet de conclure à la convergence vers 0 du saut de dualité en ce qui concerne l'algorithme au plus grand pas.

Pour terminer, intéressons-nous à la complexité de cet algorithme :

Théorème 3.3 *L'algorithme au plus grand pas calcule une solution ϵ -optimale en $O(\sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$ itérations. Ainsi, l'algorithme au plus grand pas possède la même complexité algorithmique que les deux algorithmes précédents.*

preuve : Nous savons que

$$\begin{aligned} X^{(k+1)} \bullet Z^{(k+1)} &= \gamma^{(k)} X^{(k)} \bullet Z^{(k)} \\ &\leq \gamma X^{(k)} \bullet Z^{(k)} \end{aligned}$$

où $\gamma = \frac{1}{1 + \frac{1}{\sqrt{2n}}}$ paramètre de centrage de l'algorithme à petits pas.

Ainsi la preuve du théorème 3.1 affirmant que la complexité de l'algorithme à petits pas est de l'ordre de $\sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon}$ reste valable.

Nous pouvons donc conclure que l'algorithme au plus grand pas possède la même complexité que celle de l'algorithme à petits pas.

cqfd

De nouveau dans cet algorithme, nous nous sommes restreints à un pas de longueur égale à 1 à chaque itération. L'algorithme à longs pas modifie cela.

3.5.5 Algorithme à longs pas

Tout comme l'algorithme à longs pas pour la programmation linéaire, nous choisissons plutôt que \mathcal{N}_2 le voisinage $\mathcal{N}_\infty^-(\beta)$ où β est une constante de l'intervalle $]0, 1[$ satisfaisant $\frac{1}{\beta(1-\beta)} = O(1)$. Nous définissons comme pour la programmation linéaire $t^{(k)}$ comme étant le plus grand pas possible dans la direction de Newton $(\Delta X^{(k)}, \Delta Z^{(k)})$ qui permet à l'itéré de ne pas quitter le voisinage $\mathcal{N}_\infty^-(\beta)$ pour $0 \leq t \leq t^{(k)}$. De plus, nous choisissons $\gamma^{(k)} \in]0, 1[$, le paramètre de centrage, tel que $\frac{1}{1-\gamma^{(k)}} = O(1)$ et $\frac{1}{\gamma^{(k)}} = O(1) \quad \forall k$.

Nous voyons tout d'abord que le choix de $t^{(k)}$ impose aux itérés de rester dans le voisinage.

Ensuite, nous examinons la complexité de cet algorithme qui est moins bonne que les trois algorithmes précédents. En effet,

Théorème 3.4 *L'algorithme à long pas calcule une solution ϵ -optimale en $O(\bar{n} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$ itérations.*

preuve :

Nous savons que

$$\begin{aligned} X^{(k+1)} \bullet Z^{(k+1)} &= \bar{n} \mu^{(k+1)} \\ &= \bar{n} (1 - t^{(k)} + \gamma^{(k)} t^{(k)}) \mu^{(k)} \\ &= (1 - t^{(k)} + \gamma^{(k)} t^{(k)}) X^{(k)} \bullet Z^{(k)}. \end{aligned}$$

Nous pouvons écrire $(1 - t^{(k)} + \gamma^{(k)} t^{(k)})$ comme $(1 - t^{(k)}(1 - \gamma^{(k)}))$. De plus, le lemme 3.6 affirme que si $0 < \gamma < 1, V \in \mathcal{N}_\infty^-(\beta)$ où $\beta \in]0, 1[$, alors $V(t) \in \mathcal{N}_\infty^-(\beta) \quad \forall t \geq 0$ tel que $0 \leq \bar{n}t \leq \frac{2\beta\gamma}{\frac{\beta\gamma^2}{1-\beta} + (1-\gamma^2)}$.

Comme $t^{(k)}$ est le plus grand pas possible tel que $V^{(k+1)} \in \mathcal{N}_\infty^-(\beta)$, nous avons que $t^{(k)}$ satisfait

$$\bar{n}t^{(k)} \geq \frac{2\beta\gamma^{(k)}}{\frac{\beta\gamma^{(k)2}}{1-\beta} + (1-\gamma^{(k)2})}.$$

Il suit que

$$t^{(k)}(1 - \gamma^{(k)}) \geq \frac{2\beta\gamma^{(k)}(1 - \gamma^{(k)})}{\bar{n}(\frac{\beta\gamma^{(k)2}}{1-\beta} + (1 - \gamma^{(k)2}))}.$$

Or, par hypothèse, nous savons que $\frac{1}{1-\gamma^{(k)}} \leq c_1, \frac{1}{\gamma^{(k)}} \leq c_2$ ainsi que $\frac{1}{\beta(1-\beta)} \leq c_3$.

Utilisant tout ce que nous connaissons jusqu'à présent, minorons le numérateur $2\beta\gamma^{(k)}(1 - \gamma^{(k)})$

$$2\beta\gamma^{(k)}(1 - \gamma^{(k)}) \geq \frac{2\beta}{c_1 c_2}$$

et majorons le dénominateur

$$\bar{n}(\frac{\beta\gamma^{(k)2}}{1-\beta} + (1 - \gamma^{(k)2})) \leq \frac{\bar{n}}{\beta(1-\beta)}(\beta^2\gamma^{(k)2} + \beta(1-\beta)(1 - \gamma^{(k)2}))$$

$$\begin{aligned}
&\leq \bar{n}c_3\beta(\beta\gamma^{(k)^2} + 1) \\
&\quad \text{car } (1 - \beta)(1 - \gamma^{(k)^2}) \leq 1 \\
&\leq 2\bar{n}c_3\beta \quad \text{car } \beta\gamma^{(k)^2} \leq 1
\end{aligned}$$

Donc, nous obtenons successivement

$$\begin{aligned}
1 - t^{(k)}(1 - \gamma^{(k)}) &\leq 1 - \frac{2\beta}{c_1c_2c_3\bar{n}2\beta} \\
&= 1 - \frac{1}{c_1c_2c_3\bar{n}} \\
&= 1 - \frac{c}{\bar{n}} \quad \text{où } c = \frac{1}{c_1c_2c_3}.
\end{aligned}$$

Le saut de dualité $X^{(k+1)} \bullet Z^{(k+1)}$ peut dès lors être borné comme suit :

$$\begin{aligned}
X^{(k+1)} \bullet Z^{(k+1)} &= (1 - t^{(k)}(1 - \gamma^{(k)}))X^{(k)} \bullet Z^{(k)} \\
&\leq (1 - \frac{c}{\bar{n}})X^{(k)} \bullet Z^{(k)} \\
\log(X^{(k+1)} \bullet Z^{(k+1)}) &\leq \log(1 - \frac{c}{\bar{n}}) + \log(X^{(k)} \bullet Z^{(k)}) \\
&\leq (k+1)\log(1 - \frac{c}{\bar{n}}) + \log(X^{(0)} \bullet Z^{(0)}).
\end{aligned}$$

Or, nous savons que $\log(1 + \beta) \leq \beta \quad \forall \beta \geq -1$. Dans notre cas, $-\frac{c}{\bar{n}} \geq -1$ car $c_1c_2c_3 \geq 1$, c'est-à-dire $\frac{1}{c_1c_2c_3} = c \leq 1 \leq \bar{n}$.

Par conséquent, nous avons

$$\begin{aligned}
\log(X^{(k+1)} \bullet Z^{(k+1)}) &\leq -(k+1)\frac{c}{\bar{n}} + \log(X^{(0)} \bullet Z^{(0)}) \\
X^{(k+1)} \bullet Z^{(k+1)} &\leq e^{-(k+1)\frac{c}{\bar{n}}}(X^{(0)} \bullet Z^{(0)})
\end{aligned}$$

Or, $(X^{(k+1)}, Z^{(k+1)})$ est une solution ϵ -optimale

$$\iff X^{(k+1)} \bullet Z^{(k+1)} \leq \epsilon.$$

Par cette dernière équation, nous trouvons successivement

$$\begin{aligned}
&e^{-(k+1)\frac{c}{\bar{n}}}(X^{(0)} \bullet Z^{(0)}) = \epsilon \\
\iff &-(k+1)\frac{c}{\bar{n}} = \log\left(\frac{\epsilon}{X^{(0)} \bullet Z^{(0)}}\right) \\
\iff &k+1 = \frac{\bar{n}}{c} \log\left(\frac{X^{(0)} \bullet Z^{(0)}}{\epsilon}\right) \\
\iff &k = O\left(\bar{n} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon}\right).
\end{aligned}$$

cqfd

Comme souligné plus haut, la complexité algorithmique de l'algorithme à longs pas est plus grande que celles des autres algorithmes présentés.

Nous avons donc développé quatre algorithmes de type primal-dual de points intérieurs et de suivi de chemin qui nous donne une solution primale-duale ϵ -optimale d'un problème primal-dual de programmation semi-définie, en supposant toutefois l'admissibilité forte primale et duale.

3.6 Directions de descente en programmation semi-définie pour les méthodes primales-duales de points intérieurs et de suivi de chemin

Jusqu'à présent, de nombreuses recherches ont été menées sur le choix d'une direction de descente en programmation semi-définie pour les méthodes primales-duales de points intérieurs et de suivi de chemin. Nous en donnons un très bref aperçu dans cette sixième section.

Dans un premier temps, nous nous penchons sur la recherche d'une telle direction basée sur les barrières dites "self-scaled". Nous étudions celle de Nesterov-Todd dans un cadre légèrement plus général que celui de la programmation semi-définie. Nous montrons que Nesterov et Todd furent les premiers à découvrir la direction que nous avons présentée dans ce chapitre.

Dans un second temps, nous nous intéressons aux directions trouvées en prenant une matrice particulière comme cible pour une certaine matrice caractérisant l'itéré suivant. Nous étudions spécialement la direction proposée par Monteiro et Tsuchiya.

Dans un troisième temps, nous nous attachons aux directions calculées à partir de cibles choisies pour les valeurs propres de la matrice XZ . Nous remarquons de ce fait que la direction (3.20) étudiée jusqu'à présent et celle de Monteiro-Tsuchiya font partie d'une même grande famille de directions.

3.6.1 Recherche d'une direction de descente basée sur les barrières dites "self-scaled"

Jusqu'à présent, nous avons défini la trajectoire centrale primale-duale comme l'ensemble des paires primales-duales admissibles (X, Z) pour lesquelles $XZ = \mu I$ pour un certain $\mu > 0$. De plus, nous avons calculé la direction de Newton en transformant puis linéarisant le système suivant :

$$\begin{cases} X(t)Z(t) = \mu I \\ X(t) \in (B + \mathcal{A}) \cap \mathcal{H}_n^{++} \\ Z(t) \in (C + \mathcal{A}^\perp) \cap \mathcal{H}_n^{++} \end{cases}$$

En d'autres mots, nous avons pris pour cible un point de la trajectoire centrale correspondant à une valeur de μ , point que nous appelons μ -centre.

Intéressons-nous maintenant à une autre approche du calcul de la direction : l'approche symétrique primale-duale utilisant des barrières dites "self-scaled" qui fut proposée par Nesterov et Todd dans un cadre plus restreint que celui de la programmation conique convexe mais légèrement plus général que celui de la programmation semi-définie. Nous découvrons que cette approche est une amorce à celle étudiée dans ce chapitre.

Définissons dans un premier temps le problème.

Soit K un cône convexe fermé, pointé, solide dans un espace vectoriel réel E de dimension finie et soit E^* l'espace dual correspondant.

Nous notons le produit scalaire $\langle s, x \rangle$ pour $x \in E$ et $s \in E^*$ et nous définissons le problème primal comme

$$\min \langle c, x \rangle \quad \text{s.c.} \quad Ax = b \text{ et } x \in K \quad (3.21)$$

où $b \in Y^*$, $c \in E^*$ et A est un opérateur linéaire surjectif allant de E dans Y^* .

Si nous nous basons sur l'écriture vectorielle d'un problème de programmation semi-définie, (voir 1.6), nous nous rendons compte que celui-ci peut se mettre sous la forme (3.21) en posant $\langle c, x \rangle = \sum_{i=1}^{\bar{n}} c_i x_i$, $\text{Ker} A = \mathcal{A}$ et $K = \mathcal{H}^+$.

De façon similaire à la programmation linéaire, le dual se définit alors comme :

$$\max \langle b, y \rangle \quad \text{s.c.} \quad A^*y + s = c \text{ et } s \in K^* \quad (3.22)$$

où A^* est l'adjointe de A allant de Y^* dans E^* , $y \in Y$ et $K^* = \{s \in E^* \mid \langle s, x \rangle \geq 0 \quad \forall x \in K\}$.

Dans un second temps, intéressons-nous à la caractérisation de la trajectoire centrale et du μ -centre donnée par Nesterov-Todd.

La trajectoire centrale primale $\{x(\mu) : \mu > 0\}$ pour (3.21) est définie comme suivant :

$$x(\mu) = \operatorname{argmin}_x \left\{ \frac{1}{\mu} \langle c, x \rangle + F(x) \mid Ax = b \text{ et } x \in K \right\}$$

où F est une barrière "self-scaled".

Il n'est pas nécessaire ici pour la compréhension de connaître la définition d'une barrière dite "self-scaled". Néanmoins, pour plus de renseignements, le lecteur peut consulter la définition 2.1. de [9].

La trajectoire centrale duale $\{y(\mu), s(\mu) : \mu > 0\}$ pour (3.22) est quant à elle définie par :

$$(y(\mu), s(\mu)) = \operatorname{argmin}_{y,s} \left\{ -\frac{1}{\mu} \langle b, y \rangle + F^*(s) \mid A^*y + s = c, y \in Y \text{ et } s \in K^* \right\}.$$

Par conséquent, la trajectoire centrale primale-duale pour le problème primal-dual suivant :

$$\min(\langle c, x \rangle - \langle b, y \rangle) \quad \text{s.c.} \quad Ax = b, A^*y + s = c \text{ et } x \in K, y \in Y, s \in K^*$$

est la combinaison des deux trajectoires centrales : $\{w(\mu) = (x(\mu), y(\mu), s(\mu)) : \mu > 0\}$.

Le problème étant bien posé, nous pouvons enfin nous pencher sur les deux directions proposées par Nesterov et Todd : la direction affine et la direction de centrage.

Etudions tout d'abord la direction affine.

Soient $x \in S^0(P) := \{x \in \text{int}K | Ax = b\}$, $(y, s) \in S^0(D) := \{(y, s) \in Y \times \text{int}K^* | A^*y + s = c\}$ et $z \in \text{int}K$ tel que $s = F''(z)x$ où $F''(z)$ est symétrique.

La direction affine en le point primal-dual (x, y, s) est la solution (p_x, p_y, p_s) du système linéaire suivant :

$$\begin{cases} F''(z)p_x + p_s = s \\ Ap_x = 0 \\ A^*p_y + p_s = 0. \end{cases}$$

Nous remarquons que ce système linéaire ressemble très fort à celui qui nous a permis de calculer notre direction de Newton :

$$\begin{cases} \Delta X + D(X, Z)\Delta Z D(X, Z) = \gamma\mu Z^{-1} - X \\ \Delta \dot{X} \in \mathcal{A} \text{ et } \Delta Z \in \mathcal{A}^\perp. \end{cases}$$

En effet, $F''(z)p_x + p_s = s$ peut se réécrire $p_x + (F''(z))^{-1}p_s = x$. En prenant $\gamma = 0$ dans le système ci-dessus, ce qui revient en fait à choisir une direction "prédicteur", nous avons que $\Delta X + D(X, Z)\Delta Z D(X, Z) = -X$ est assez semblable à $p_x + (F''(z))^{-1}p_s = x$. $D(X, Z)$ et $(F''(z))^{-1}$ jouent le même rôle.

De plus, $Ap_x = 0$ et $A^*p_y + p_s = 0$ nous permettent de déduire que p_s et p_x sont orthogonales comme ΔX et ΔZ le sont. En effet, $\langle p_s, p_x \rangle = \langle -A^*p_y, p_x \rangle = \langle -p_y, Ap_x \rangle = \langle -p_y, 0 \rangle = 0$.

Une dernière similitude réside dans le saut de dualité. Nous avons $X^{(k+1)} \bullet Z^{(k+1)} = (1 - t^{(k)})X^{(k)} \bullet Z^{(k)}$. De nouveau, en posant α la longueur de pas ainsi que $x(\alpha) := x - \alpha p_x$, $y(\alpha) := y - \alpha p_y$ et $s(\alpha) := s - \alpha p_s$, nous obtenons successivement

$$\begin{aligned} \langle s(\alpha), x(\alpha) \rangle &= \langle s - \alpha p_s, x - \alpha p_x \rangle \\ &= \langle s, x \rangle - \alpha \langle p_s, x \rangle - \alpha \langle s, p_x \rangle + \alpha^2 \langle p_s, p_x \rangle \\ &= \langle s, x \rangle - \alpha (\langle p_s, x \rangle + \langle s, p_x \rangle) \\ &= \langle s, x \rangle - \alpha (\langle p_s, x \rangle + \langle (F''(z))^{-1}s, F''(z)p_x \rangle) \\ &= \langle s, x \rangle - \alpha (\langle p_s, x \rangle + \langle x, s - p_s \rangle) \\ &= \langle s, x \rangle (1 - \alpha). \end{aligned}$$

Ensuite, étudions la direction de centrage.

Soient $x \in S^0(P) := \{x \in \text{int}K | Ax = b\}$, $(y, s) \in S^0(D) := \{(y, s) \in Y \times \text{int}K^* | A^*y + s = c\}$ et $z \in \text{int}K$ tel que $s = F''(z)x$ où $F''(z)$ est symétrique.

La direction de centrage en le point primal-dual (x, y, s) est la solution (d_x, d_y, d_s) du système linéaire suivant :

$$\begin{cases} F''(z)d_x + d_s = s + \mu(x, s)F'(x) \\ Ad_x = 0 \\ A^*d_y + d_s = 0 \end{cases}$$

où $\mu(x, s) = \frac{1}{\nu} < x, s >$ et ν est une caractéristique de F . Nous remarquons que ce système linéaire ressemble très fort à celui qui nous a permis de calculer notre direction de Newton :

$$\begin{cases} \Delta X + D(X, Z)\Delta Z D(X, Z) = \gamma\mu Z^{-1} - X \\ \Delta X \in \mathcal{A} \text{ et } \Delta Z \in \mathcal{A}^\perp \end{cases}$$

En effet, $F'''(z)d_x + d_s = s + \mu(x, s)F'(x)$ peut se réécrire $d_x + (F'''(z))^{-1}d_s = x + \mu(x, s)F'_*(s)$, par [9]. En prenant $\gamma = 1$ dans le système ci-dessus, ce qui revient en fait à choisir une direction "correcteur", nous avons que $\Delta X + D(X, Z)\Delta Z D(X, Z) = \mu Z^{-1} - X$ est assez semblable à $d_x + (F'''(z))^{-1}d_s = x + \mu(x, s)F'_*(s)$. $D(X, Z)$ et $(F'''(z))^{-1}$ jouent le même rôle. De plus, $Ad_x = 0$ et $A^*d_y + d_s = 0$ nous permettent de déduire l'orthogonalité de d_s et d_x tout comme ΔX et ΔZ le sont.

Par cette théorie, Nesterov et Todd furent les premiers à obtenir la direction de Newton $(\Delta X, \Delta Z)$ définie en (3.20) qui est nommée souvent direction de "Nesterov-Todd".

3.6.2 Recherche d'une direction de descente basée sur les cibles matricielles

La direction de descente pour une méthode primale-duale de points intérieurs de suivi de chemin peut être calculée en se basant sur un système linéaire qui impose entre autres à une certaine matrice de s'approcher le plus possible d'une autre.

La direction de Monteiro-Tsuchiya fait partie de cette famille de directions-là.

Au début de ce chapitre, nous avons cherché dans l'espace initial notre direction de Newton $(\Delta X, \Delta Z)$ à partir d'un point (X, Z) en nous basant sur le système suivant :

$$\begin{cases} (X + \Delta X)(Z + \Delta Z) = \gamma\mu I \\ \Delta X \in \mathcal{A} \quad \Delta Z \in \mathcal{A}^\perp \end{cases}$$

Nous l'avons exprimé dans l'espace transformé puis linéarisé.

Monteiro et Tsuchiya ne sont pas passés par l'espace transformé pour trouver leur direction et ont donc proposé de linéariser ce système après une légère modification.

En effet, nous écrivons les équivalences suivantes :

$$\begin{aligned} & (X + \Delta X)(Z + \Delta Z) = \gamma\mu I \\ \iff & L^H(X + \Delta X)(Z + \Delta Z)L^{-H} = \gamma\mu I \\ \iff & L^{-1}(Z + \Delta Z)L^{-H} = \gamma\mu(L^H(X + \Delta X)L)^{-1} \\ \iff & (L^H(X + \Delta X)L)^{\frac{1}{2}}L^{-1}(Z + \Delta Z)L^{-H}(L^H(X + \Delta X)L)^{\frac{1}{2}} = \gamma\mu I. \end{aligned}$$

La méthode de Monteiro-Tsuchiya consiste donc à linéariser le système

$$\begin{cases} R_X(1)L^{-1}(Z + \Delta Z)L^{-H}R_X(1) = \gamma\mu I \\ R_X(t)^2 = L^H(X + t\Delta X)L \\ \Delta X \in \mathcal{A} \quad \Delta Z \in \mathcal{A}^\perp \end{cases}$$

où L est une matrice inversible et la matrice hermitienne $R_X(t) = (L^H(X + t\Delta X)L)^{\frac{1}{2}}$.
 Pour simplifier encore quelque peu ce système, nous posons $L_X(t) = L^{-H}R_X(t)$ et obtenons

$$\begin{cases} L_X(1)^H(Z + \Delta Z)L_X(1) = \gamma\mu I \\ L_X(t)L_X(t)^H = X + t\Delta X \\ L^H L_X(t) \in \mathcal{H} \\ \Delta X \in \mathcal{A} \\ \Delta Z \in \mathcal{A}^\perp. \end{cases} \quad (3.23)$$

La première équation de ce système traduit le fait que nous prenons $\gamma\mu I$ comme cible pour la matrice $L_X(1)^H(Z + \Delta Z)L_X(1)$. En fait, posant $W(t) := L_X(t)^H(Z + t\Delta Z)L_X(t)$, cette équation devient $W(1) = \gamma\mu I$ très semblable à celle que nous avons précédemment : $V(1)^2 = \gamma\mu I$.

Pour obtenir la direction recherchée, linéarisons le système (3.23).
 Intéressons-nous tout d'abord à notre première équation. Nous avons $W(1) = \gamma\mu I = W(0) + W^{(1)}(0)$ où $W(0) = L_X^H Z L_X$ avec $L_X := L_X(0)$. Par ailleurs, nous avons

$$(X + t\Delta X)(Z + t\Delta Z) = L_X(t)W(t)L_X(t)^{-1}. \quad (3.24)$$

En effet, nous écrivons

$$\begin{aligned} L_X(t)L_X(t)^H &= L^{-H}R_X(t)R_X(t)^H L^{-1} \\ &= L^{-H}R_X(t)^2 L^{-1} \quad \text{car } R_X(t) \in \mathcal{H} \\ &= (X + t\Delta X). \end{aligned}$$

Il suit que

$$\begin{aligned} L_X(t)W(t)L_X(t)^{-1} &= L_X(t)L_X(t)^H(Z + t\Delta Z)L_X(t)L_X(t)^{-1} \\ &= (X + t\Delta X)(Z + t\Delta Z). \end{aligned}$$

Effectuant la multiplication, nous obtenons

$$L_X(t)W(t)L_X(t)^{-1} = XZ + tX\Delta Z + t(\Delta X)Z + t^2\Delta X\Delta Z.$$

Dérivons par rapport à t :

$$\begin{aligned} X\Delta Z + (\Delta X)Z + 2t\Delta X\Delta Z &= L_X^{(1)}(t)W(t)L_X^{-1}(t) + L_X(t)W^{(1)}(t)L_X^{-1}(t) \\ &\quad - L_X(t)W(t)L_X(t)^{-1}L_X^{(1)}(t)L_X^{-1}(t) \end{aligned}$$

par (3.8). Considérant cette expression en $t = 0$, nous avons, puisque $W^{(1)}(0) = \gamma\mu I - W(0)$,

$$\begin{aligned} X\Delta Z + (\Delta X)Z &= L_X^{(1)}(0)W(0)L_X^{-1} + L_X(\gamma\mu I - W(0))L_X^{-1} - L_XW(0)L_X^{-1}L_X^{(1)}(0)L_X^{-1} \\ &= \gamma\mu I - XZ + L_X^{(1)}(0)L_X^{-1}XZ - XZL_X^{(1)}(0)L_X^{-1} \end{aligned}$$

car $L_X W(0) L_X^{-1} = XZ$ par (3.24).

La seconde équation de (3.23) qui est $L_X(t) L_X(t)^H = X + t\Delta X$ devient, après linéarisation de $L_X(t)$ $2P_H(L_X^{(1)}(0) L_X^H) = \Delta X$. En effet, nous obtenons successivement

$$\begin{aligned} (L_X + tL_X^{(1)}(0))(L_X + tL_X^{(1)}(0))^H &= X + t\Delta X \\ L_X L_X^H + tL_X L_X^{(1)}(0)^H + tL_X^{(1)}(0) L_X^H + t^2 L_X^{(1)}(0) L_X^{(1)}(0)^H &= L_X L_X^H + t\Delta X \\ t(L_X L_X^{(1)}(0)^H + L_X^{(1)}(0) L_X^H) &= t\Delta X \end{aligned}$$

qui est équivalent à $2P_H(L_X^{(1)}(0) L_X^H) = \Delta X$.

De même la troisième équation de (3.23) qui est $L^H L_X(t) \in \mathcal{H}$ se transforme en les deux suivantes grâce à la linéarisation de $L_X(t)$:

$$L^H L_X \in \mathcal{H}, \quad L^H L_X^{(1)}(0) \in \mathcal{H}.$$

Il suit de tout ceci que les directions de Monteiro-Tsuchiya sont solutions du système :

$$\begin{cases} X\Delta Z + (\Delta X)Z = \gamma\mu I - XZ + L_X^{(1)}(0) L_X^{-1} XZ - XZ L_X^{(1)}(0) L_X^{-1} \\ 2P_H(L_X^{(1)}(0) L_X^H) = \Delta X \\ L^H L_X \in \mathcal{H} \quad L^H L_X^{(1)}(0) \in \mathcal{H} \\ \Delta X \in \mathcal{A} \\ \Delta Z \in \mathcal{A}^\perp. \end{cases} \quad (3.25)$$

3.6.3 Recherche d'une direction de descente basée sur les cibles pour les valeurs propres de XZ

En ce qui concerne le calcul de la direction $(\Delta X, \Delta Z)$, après des méthodes se basant sur des barrières dites "self-scaled" et les nombreuses variantes possibles usant de cibles matricielles, la troisième grande catégorie de méthodes est celle des approches qui considèrent une cible pour les valeurs propres de la matrice XZ .

La méthode que nous avons développée dans la section 3.2. appartient à cette catégorie. En effet, nous avons pris $\gamma\mu I$ comme cible de $V^2(1)$ où $V^2(1)$ est la matrice diagonale des valeurs propres de $X(1)Z(1)$. Cependant, rappelons qu'alors nous avons effectué le choix spécifique de $G(t) = D(X(t), Z(t))^\frac{1}{2}$ pour $G(t)$, matrice de transformation vers l'espace transformé. De cette manière, nous avons obtenu une matrice $G(t)$ hermitienne pour tout t , ce qui nous a grandement simplifié nos calculs dans la recherche de la direction $(\Delta X, \Delta Z)$.

Nous allons montrer tout d'abord ce qu'il advient de la direction de Newton pour un $G(t)$ non spécifié mais satisfaisant toujours le minimum de contraintes exigées, c'est-à-dire $G(t)G(t)^H = D(X(t), Z(t))$ ainsi que $G(0) = I$. En effet, toutes ces définitions de $G(t)$ conduisent à des directions de Newton différentes. Ensuite, nous allons voir quelle définition

associer à $G(t)$ pour retrouver la direction de Monteiro-Tsuchiya.

Dans la section 3.2., nous avons calculé la direction de Newton $(\Delta X, \Delta Z)$, associée à un point (X, Z) où $X \in (B + \mathcal{A}) \cap \mathcal{H}_{\bar{n}}^{++}$ et $Z \in (C + \mathcal{A}^\perp) \cap \mathcal{H}_{\bar{n}}^{++}$ dans l'espace initial. Pour ce faire, nous avons tout d'abord cherché une direction de Newton, (D_X, D_Z) , dans l'espace transformé. En fait, nous étions amenés à linéariser le système suivant :

$$\begin{cases} V^2(1) = \gamma\mu I \\ D_X \in \mathcal{A}(L_d) \\ D_Z \in \mathcal{A}^\perp(L_d). \end{cases} \quad (3.26)$$

Examinant ce système non-linéaire, nous constatons tout d'abord que la première équation de ce système exige en fait que $\gamma\mu$ soit une cible pour chaque valeur propre de $X(1)Z(1)$. Or, à ce stade-ci, nous voyons que, quelle que soit la définition de $G(t)$, la direction de Newton est toujours solution de ce système mais linéarisé. Par conséquent, quelle que soit la définition de $G(t)$, c'est-à-dire quelle que soit la direction de Newton choisie dans cette section, les directions calculées ci-dessous correspondent à des méthodes qui prennent une cible pour les valeurs propres de la matrice XZ .

Dans notre raisonnement de la section 3.2 concernant la recherche de la direction de Newton, nous sommes arrivés, sans avoir posé $G(t) = D(X(t), Z(t))^{\frac{1}{2}}$, à l'équation (3.9) :

$$V^{(1)}(0) = \frac{1}{2}(Z^{(1)}(0) + X^{(1)}(0)) + \mathcal{P}_{\mathcal{H}}((G^{(1)}(0)^H - G^{(1)}(0))V).$$

A ce moment précis, nous avons considéré $G(t) = D(X(t), Z(t))^{\frac{1}{2}} \in \mathcal{H}$ et par conséquent annulé $\mathcal{P}_{\mathcal{H}}((G^{(1)}(0)^H - G^{(1)}(0))V)$. Si nous n'effectuons pas ce choix, $G(t)$ n'est alors pas nécessairement hermitienne. La simplification n'est donc plus possible. Examinons cette éventualité menant à plusieurs directions.

Posant $\mathcal{P}_{\mathcal{H}^\perp}G^{(1)}(0) = \tilde{D}_G$, nous obtenons $V^{(1)}(0) = \frac{1}{2}(Z^{(1)}(0) + \bar{X}^{(1)}(0)) - 2\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V)$ car, par définition, nous avons $\tilde{D}_G = \frac{G^{(1)}(0) - G^{(1)}(0)^H}{2}$.

Or, nous savons que $D_Z = Z^{(1)}(0)$ et $D_X = \bar{X}^{(1)}(0)$ par (3.6). Il suit que $V^{(1)}(0) = \frac{1}{2}(D_X + D_Z) - 2\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V)$.

Or, linéariser l'équation $V^2(1) = \gamma\mu I$ revient donc à faire

$$\gamma\mu I = V(1)^2 \approx V(0)^2 + \frac{dV(t)^2}{dt}\bigg|_{t=0.1}$$

c'est-à-dire,

$$\begin{aligned} \gamma\mu I &= V^2 + 2\mathcal{P}_{\mathcal{H}}(VV^{(1)}(0)) \\ &= V^2 + 2\mathcal{P}_{\mathcal{H}}(V(\frac{1}{2}(D_X + D_Z) - 2\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V))) \\ &= V^2 + \mathcal{P}_{\mathcal{H}}(V(D_X + D_Z)) - 4\mathcal{P}_{\mathcal{H}}(V\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V)). \end{aligned}$$

Le système linéarisé donnant la direction de Newton (D_X, D_Z) est donc

$$\begin{cases} \mathcal{P}_{\mathcal{H}}(V(D_X + D_Z)) = \gamma\mu I - V^2 + 4\mathcal{P}_{\mathcal{H}}(V\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V)) \\ D_X \in \mathcal{A}(L_d) \\ D_Z \in \mathcal{A}^{\perp}(L_d) \\ \tilde{D}_G \in \mathcal{H}^{\perp} \end{cases} \quad (3.27)$$

Pour résoudre la première équation, nous nous aidons de nouveau du lemme 3.2 qui dit qu'étant donné V une matrice diagonale à éléments positifs et Y une matrice hermitienne du même ordre, alors

$$\mathcal{P}_{\mathcal{H}}(VX) = Y \iff X_{ij} = \frac{2Y_{ij}}{V_{ii} + V_{jj}} \quad \forall i, j.$$

En ce qui nous concerne, nous avons V une matrice diagonale positive (voir 3.1.2.). Nous voulons $D_X + D_Z$ hermitienne. Par conséquent, la solution de

$$\mathcal{P}_{\mathcal{H}}(V(D_X + D_Z)) = \gamma\mu I - V^2 + 4\mathcal{P}_{\mathcal{H}}(V\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V))$$

est

$$D_X + D_Z = \gamma\mu V^{-1} - V + 4\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V).$$

En effet, le lemme 2.3 dit que

$$(D_X + D_Z)_{ij} = \frac{2((\gamma\mu\delta_{ij} - V_{ij}^2) + 4(\mathcal{P}_{\mathcal{H}}(V\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V)))_{ij})}{V_{ii} + V_{jj}}$$

– Pour le cas où $i \neq j$:

$$(D_X + D_Z)_{ij} = \frac{8(\mathcal{P}_{\mathcal{H}}(V\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V)))_{ij}}{V_{ii} + V_{jj}}.$$

Or, nous écrivons

$$\begin{aligned} (\mathcal{P}_{\mathcal{H}}(V\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V)))_{ij} &= \frac{1}{2}(((V\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V))_{ij} + (\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V)V)_{ij}) \\ &= \frac{1}{2}(V_{ii}(\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V))_{ij} + (\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V))_{ij}V_{jj}) \\ &= \frac{1}{2}(\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V))_{ij}(V_{ii} + V_{jj}). \end{aligned}$$

$$\text{Or } (\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V))_{ij} = (\tilde{D}_{Gij}V_{jj} + V_{ii}\tilde{D}_{Gij}^H)\frac{1}{2}.$$

$$\text{Donc } (\mathcal{P}_{\mathcal{H}}(V\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V)))_{ij} = \frac{1}{4}(\tilde{D}_{Gij}V_{jj} + V_{ii}\tilde{D}_{Gij}^H)(V_{ii} + V_{jj}).$$

$$\text{Par conséquent, } (D_X + D_Z)_{ij} = 2(\tilde{D}_{Gij}V_{jj} + V_{ii}\tilde{D}_{Gij}^H) = 4(\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V))_{ij}.$$

– Pour le cas où $i=j$:

$$\begin{aligned}
(D_X + D_Z)_{ii} &= \frac{2((\gamma\mu - V_{ii}^2) + 4(\mathcal{P}_{\mathcal{H}}(V\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V)))_{ii})}{2V_{ii}} \\
&= \gamma\mu V_{ii}^{-1} - V_{ii} + 4 \frac{1}{4} \frac{(\tilde{D}_{Gii} + \tilde{D}_{Gii}^H) V_{ii} 2V_{ii}}{V_{ii}} \\
&= \gamma\mu V_{ii}^{-1} - V_{ii} + 2(\tilde{D}_{Gii} + \tilde{D}_{Gii}^H) V_{ii} \\
&= \gamma\mu V_{ii}^{-1} - V_{ii} + 4(\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V))_{ii}.
\end{aligned}$$

Utilisant le fait que

$$\begin{aligned}
\mathcal{P}_{\mathcal{H}^\perp} G^{(1)}(0) = \tilde{D}_G &= \frac{G^{(1)}(0) - G^{(1)}(0)^H}{2} \\
&= -\frac{G^{(1)}(0)^H - G^{(1)}(0)}{2} \\
&= -\tilde{D}_G^H,
\end{aligned}$$

Nous observons que

$$\begin{aligned}
4(\mathcal{P}_{\mathcal{H}}(\tilde{D}_G V)) &= 4 \frac{\tilde{D}_G V + V^H \tilde{D}_G^H}{2} \\
&= 2(\tilde{D}_G V - V \tilde{D}_G).
\end{aligned}$$

En conclusion, la direction de Newton (D_X, D_Z) au point (\bar{X}, \bar{Z}) dans l'espace transformé satisfait

$$D_X + D_Z = \gamma\mu V^{-1} - V + 2(\tilde{D}_G V - V \tilde{D}_G)$$

ainsi que

$$D_X \in \mathcal{A}(L_d), \quad D_Z \in \mathcal{A}^\perp(L_d), \quad \tilde{D}_G \in \mathcal{H}^\perp.$$

Tirons quelques résultats utiles.

D'une part,

$$\begin{aligned}
(D_X + D_Z)_{ij} &= (\gamma\mu V^{-1} - V)_{ij} + 2(\tilde{D}_G V - V \tilde{D}_G)_{ij} \\
&= (\gamma\mu V^{-1} - V)_{ij} + 2\tilde{D}_{Gij}(V_{jj} - V_{ii}).
\end{aligned} \tag{3.28}$$

Nous remarquons ainsi que le choix de la direction de ce chapitre spécifié en (3.11) consiste en l'annulation de \tilde{D}_G , c'est-à-dire en l'annulation du second terme de la somme expliquant $(D_X + D_Z)_{ij}$.

D'autre part,

$$\begin{aligned}
\|D_X + D_Z\|_F^2 &= \|\gamma\mu V^{-1} - V + 2(\tilde{D}_G V - V \tilde{D}_G)\|_F^2 \\
&= \|\gamma\mu V^{-1} - V\|_F^2 + 4\|(\tilde{D}_G V - V \tilde{D}_G)\|_F^2 \\
&\quad + 4((\gamma\mu V^{-1} - V) \bullet (\tilde{D}_G V - V \tilde{D}_G)).
\end{aligned}$$

Or, nous avons

$$(\mu V^{-1} - V) \bullet (\tilde{D}_G V - V \tilde{D}_G) = 0$$

car $\mu V^{-1} - V$ est une matrice diagonale et $\tilde{D}_G V - V \tilde{D}_G$ est une matrice ayant tous ses éléments diagonaux nuls.

Il suit que $\|D_X + D_Z\|_F^2 = \|\gamma\mu V^{-1} - V\|_F^2 + 4\|(\tilde{D}_G V - V \tilde{D}_G)\|_F^2$.

Nous remarquons ainsi que le choix effectué pour la direction dans ce chapitre caractérisée par (3.11) consiste en la direction de plus petite longueur.

Terminons par la recherche de la matrice $G(t)$ caractérisée par la matrice \tilde{D}_G nous permettant de définir la direction de Monteiro-Tsuchiya comme une direction basée sur les cibles pour les valeurs propres de XZ et calculée dans le V -espace. Dans ce but, posons $\hat{U}_X := L_d^{-1}(L_X^{(1)}(0)L_X^{-1})L_d$. Nous voyons donc que la direction de Monteiro-Tsuchiya (D_X, D_Z) satisfait

$$VD_Z + D_X V = (\gamma\mu I - V^2) + \hat{U}_X V^2 - V^2 \hat{U}_X. \quad (3.29)$$

En effet, nous avons la première équation du système (3.25) qui est

$$X\Delta Z + (\Delta X)Z = \gamma\mu I - XZ + L_X^{(1)}(0)L_X^{-1}XZ - XZL_X^{(1)}(0)L_X^{-1}$$

c'est-à-dire

$$\begin{aligned} & L_d^{-1}XL_d^{-H}L_d^H\Delta ZL_d + L_d^{-1}(\Delta X)L_d^{-H}L_d^HZL_d \\ &= \gamma\mu I - L_d^{-1}XZL_d + L_d^{-1}L_X^{(1)}(0)L_X^{-1}XZL_d - L_d^{-1}XZL_X^{(1)}(0)L_X^{-1}L_d. \end{aligned}$$

Or, puisque $L_d^{-1}XL_d^{-H} = V = L_d^HZL_d$, $L_d^H\Delta ZL_d = D_Z$, $L_d^{-1}\Delta XL_d^{-H} = D_X$ et $\hat{U}_X = L_d^{-1}(L_X^{(1)}(0)L_X^{-1})L_d$, nous retrouvons (3.29).

Désirant nous ramener à une équation du type (3.28), nous prenons la partie hermitienne de (3.29) et obtenons

$$\mathcal{P}_{\mathcal{H}}(V(D_X + D_Z)) = (\gamma\mu I - V^2) + \mathcal{P}_{\mathcal{H}}((\hat{U}_X - \hat{U}_X^H)V^2).$$

Pour résoudre cette équation, nous nous aidons de nouveau du lemme 3.2 qui dit qu'étant donnés V une matrice diagonale à éléments positifs et Y une matrice hermitienne du même ordre, alors

$$\mathcal{P}_{\mathcal{H}}(VX) = Y \iff X_{ij} = \frac{2Y_{ij}}{V_{ii} + V_{jj}} \quad \forall i, j.$$

Par conséquent, nous obtenons

$$(D_X + D_Z)_{ij} = \frac{2}{V_{ii} + V_{jj}}((\gamma\mu I - V^2)_{ij} + (\mathcal{P}_{\mathcal{H}}((\hat{U}_X - \hat{U}_X^H)V^2))_{ij}).$$

Or, nous avons successivement

$$\begin{aligned}
\mathcal{P}_{\mathcal{H}}((\hat{U}_X - \hat{U}_X^H)V^2) &= \frac{1}{2}((\hat{U}_X - \hat{U}_X^H)V^2 + V^2(\hat{U}_X^H - \hat{U}_X)) \\
&= \frac{1}{2}((U_{Xij} - U_{Xij}^H)V_{jj}^2 + V_{ii}^2(U_{Xij}^H - U_{Xij})) \\
&= \frac{1}{2}((U_{Xij} - U_{Xij}^H)(V_{jj}^2 - V_{ii}^2)).
\end{aligned}$$

Il suit que

$$(D_X + D_Z)_{ij} = (\gamma\mu V^{-1} - V)_{ij} + (V_{jj} - V_{ii})(U_{Xij} - U_{Xij}^H).$$

Nous référant à (3.28), équation générale exprimant les composantes de $D_X + D_Z$, nous déduisons que la direction de Monteiro-Tsuchiya est une direction basée sur les cibles pour les valeurs propres de XZ caractérisée par $\tilde{D}_G = \mathcal{P}_{\mathcal{H}}(\hat{U}_X)$, car $2\tilde{D}_{Gij} = (U_{Xij} - U_{Xij}^H)$.

Chapitre 4

Analyse de la convergence de la méthode primale-duale de points intérieurs et de suivi de chemin pour la programmation semi-définie

Tout comme dans le chapitre 3, considérons le problème primal de programmation semi-définie suivant :

$$\inf\{C \bullet X | X \in (B + \mathcal{A}) \cap \mathcal{H}_n^+\} \quad (4.1)$$

et son problème dual

$$\inf\{B \bullet Z | Z \in (C + \mathcal{A}^\perp) \cap \mathcal{H}_n^+\}. \quad (4.2)$$

Supposons également l'admissibilité forte primale et duale.

Jusqu'à présent, nous savons que tout algorithme, basé sur l'algorithme général développé en 3.5.1 pour la résolution de (4.1) par une méthode de points intérieurs primale-duale de suivi de chemin, est gouverné par le saut de dualité des points $(X^{(k)}, Z^{(k)})$ engendrés : $X^{(k)} \bullet Z^{(k)} = \bar{n}\mu^{(k)}$. En effet, les itérés sont générés de manière à faire tendre celui-ci vers 0. Ceci permet de conclure à la convergence des points de l'algorithme vers une solution optimale. (voir 3.1.1).

Dans ce chapitre-ci, nous nous intéressons de façon plus précise à la convergence des points $(X^{(k)}, Z^{(k)})$. Nous déterminons la solution optimale vers laquelle ceux-ci convergent et la vitesse de convergence associée. Dans ce but, nous analysons dans un premier temps la convergence des μ -centres, points se trouvant sur la trajectoire centrale, puisque les itérés n'en sont pas très éloignés. Nous prouvons que les μ -centres convergent vers (X^*, Z^*) où X^* est le centre analytique de \mathcal{F}_p^* et Z^* est celui de \mathcal{F}_d^* , à une vitesse d'ordre μ . Dans un second temps, nous utilisons cet important résultat pour démontrer la convergence super-linéaire des itérés $(X^{(k)}, Z^{(k)})$ d'un algorithme de type prédicteur-correcteur vers le couple des centres analytiques de \mathcal{F}_p^* et de \mathcal{F}_d^* .

4.1 Centre analytique de l'ensemble des solutions primales-duales optimales

Puisque le couple des centres analytiques de $\mathcal{F}_{\mathcal{P}}^*$ et de $\mathcal{F}_{\mathcal{D}}^*$ est le point de convergence des couples $(X^{(k)}, Z^{(k)})$ des algorithmes décrits dans le chapitre 3, nous sommes amenés à le définir.

En fait, nous supposons l'existence d'une solution primale-duale optimale strictement complémentaire. Celle-ci nous permet de définir une partition particulière de toute matrice hermitienne. Cette partition nous amène ensuite à définir et caractériser le couple des centres analytiques de $\mathcal{F}_{\mathcal{P}}^*$ et de $\mathcal{F}_{\mathcal{D}}^*$ par un système d'équations.

4.1.1 Partition d'une matrice hermitienne

Faisons l'hypothèse suivante :

il existe une solution primale-duale optimale strictement complémentaire pour (4.1) et (4.2). Ceci veut dire qu'il existe $(X^*, Z^*) \in \mathcal{F}_{\mathcal{P}}^* \times \mathcal{F}_{\mathcal{D}}^*$ tel que

$$\begin{cases} X^* Z^* = 0 \\ X^* + Z^* \succ 0. \end{cases}$$

Désignons par $\nu_1, \dots, \nu_k, \nu_{k+1}, \dots, \nu_{\bar{n}}$ les valeurs propres de X^* , et par $\mu_1, \dots, \mu_k, \mu_{k+1}, \dots, \mu_{\bar{n}}$ les valeurs propres de Z^* . Supposons $\nu_1, \dots, \nu_k \neq 0$ et $\nu_{k+1}, \dots, \nu_{\bar{n}} = 0$. Ainsi, nous avons $\dim \ker X^* = n - k$. L'hypothèse $X^* Z^* = 0$ entraîne que si, pour $i = 1 \dots k$, p_i est un vecteur propre de ν_i , alors p_i est un vecteur propre associé à μ_i avec μ_i nul. Nous avons $\dim \ker Z^* \geq k$. De plus, nous savons que $X^* + Z^*$ est inversible. Il suit que $\ker X^* \cap \ker Z^* = \{0\}$. Nous obtenons $\dim \ker Z^* \leq n - \dim \ker X^* = k$. Par conséquent, nous concluons que $\dim \ker Z^* = k$ et donc que $\mu_{k+1} \neq 0, \dots, \mu_{\bar{n}} \neq 0$. Si nous notons $p_{k+1}, \dots, p_{\bar{n}}$ les vecteurs propres associés à $\mu_{k+1}, \dots, \mu_{\bar{n}}$, nous avons que $p_{k+1}, \dots, p_{\bar{n}}$ sont également les vecteurs propres associés $\nu_{k+1}, \dots, \nu_{\bar{n}}$, valeurs propres nulles de X^* . Posant $P = (p_1, \dots, p_{\bar{n}})$, nous obtenons que

$$P^{-1} X^* P = \begin{pmatrix} \Lambda_B & 0 \\ 0 & 0 \end{pmatrix} \quad \text{et} \quad P^{-1} Z^* P = \begin{pmatrix} 0 & 0 \\ 0 & \Lambda_N \end{pmatrix}$$

où $\Lambda_B = \text{diag}(\nu_1 \dots \nu_k)$ et $\Lambda_N = \text{diag}(\mu_{k+1} \dots \mu_{\bar{n}})$ avec $0 \leq k \leq \bar{n}$. Finalement, comme $X^* + Z^* \succ 0$, nous avons $\nu_1 \dots \nu_k$ et $\mu_{k+1} \dots \mu_{\bar{n}}$ strictement positifs.

C'est pourquoi, nous supposons sans perdre de généralité, c'est-à-dire en appliquant une transformation unitaire aux données du problème primal-dual initial si nécessaire, que

$$X^* = \begin{pmatrix} \Lambda_B & 0 \\ 0 & 0 \end{pmatrix} \quad \text{et} \quad Z^* = \begin{pmatrix} 0 & 0 \\ 0 & \Lambda_N \end{pmatrix}$$

où $\Lambda_B = \text{diag}(\lambda_1 \dots \lambda_k)$ et $\Lambda_N = \text{diag}(\lambda_{k+1} \dots \lambda_{\bar{n}})$ avec $0 \leq k \leq \bar{n}$ et $\lambda_i > 0$ pour $1 \leq i \leq \bar{n}$.

Sur base de ces deux matrices, nous partitionnons toute matrice hermitienne X de $C^{\bar{n} \times \bar{n}}$ de la façon suivante :

$$X = \begin{pmatrix} X_B & X_U \\ X_U^H & X_N \end{pmatrix}$$

où $X_B \in C^{k \times k}$, $X_U \in C^{k \times (\bar{n}-k)}$ et $X_N \in C^{(\bar{n}-k) \times (\bar{n}-k)}$.

Avec cette nouvelle notation, nous obtenons

$$\mathcal{F}_{\mathcal{P}}^* := \{X \in \mathcal{F}_{\mathcal{P}} | C \bullet X = p^*\} \quad (4.3)$$

$$= \{X \in \mathcal{F}_{\mathcal{P}} | X_U = 0 \text{ et } X_N = 0\} \quad (4.4)$$

En effet, si nous supposons $X \in \mathcal{F}_{\mathcal{P}}^*$, nous obtenons $X \bullet Z^* = C \bullet X + B \bullet Z^* = p^* + d^* = 0$ par le théorème de Slater (2.6).

En partitionnant X et Z^* , nous avons :

$$\begin{pmatrix} X_B & X_U \\ X_U^H & X_N \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \Lambda_N \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

et par conséquent $X_U = 0$ et $X_N = 0$.

Pour démontrer l'autre inclusion, nous supposons $X \in \mathcal{F}_{\mathcal{P}}$ tel que $X_U = 0$ et $X_N = 0$. En partitionnant X et Z^* , nous avons

$$XZ^* = \begin{pmatrix} X_B & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \Lambda_N \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

et donc $X \in \mathcal{F}_{\mathcal{P}}^*$.

Par un raisonnement similaire, l'ensemble des solutions optimales duales s'écrit successivement

$$\mathcal{F}_{\mathcal{D}}^* := \{Z \in \mathcal{F}_{\mathcal{D}} | B \bullet Z = d^*\} \quad (4.5)$$

$$= \{Z \in \mathcal{F}_{\mathcal{D}} | Z_U = 0 \text{ et } Z_B = 0\}. \quad (4.6)$$

4.1.2 μ -centre et centre analytique

Puisque le premier but de ce chapitre est la preuve de la convergence des μ -centres vers le centre analytique de l'ensemble des solutions optimales primales-duales, nous rappelons la définition de μ -centre, et nous définissons le centre analytique de $\mathcal{F}_{\mathcal{P}}^*$ ainsi que celui de $\mathcal{F}_{\mathcal{D}}^*$, grâce à la partition d'une matrice hermitienne développée ci-dessus.

Etant donné $\mu \in R^{++}$, le couple $(X, Z) \in \mathcal{F}_{\mathcal{P}} \times \mathcal{F}_{\mathcal{D}}$ est le μ -centre $(X(\mu), Z(\mu)) \iff XZ = X(\mu)Z(\mu) = \mu I$.

Faisant référence à la définition de la trajectoire centrale, nous pouvons dire que la trajectoire centrale est la courbe $\{(X(\mu), Z(\mu)) | \mu > 0\}$.

Définissons maintenant un objet important de ce chapitre, le couple (X^a, Z^a) formé des centres analytiques des ensembles de solutions optimales primales et duales relatifs à nos problèmes (4.1) et (4.2).

Préoccupons-nous tout d'abord de définir le centre analytique X^a de \mathcal{F}_P^* . En vue de cet objectif, définissons le sous-espace linéaire

$$\mathcal{A}_B := \{X_B \in \mathcal{H}^{(k)} \mid \begin{pmatrix} X_B & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{A}\}. \quad (4.7)$$

L'ensemble des solutions optimales primales, \mathcal{F}_P^* , devient

$$\{X \in \mathcal{H}^{(\bar{n})} \mid X_B \in (\Lambda_B + \mathcal{A}_B) \cap \mathcal{H}_{(k)}^+, \quad X_U = 0 \quad \text{et} \quad X_N = 0\},$$

grâce à (4.4) et au fait que

$$X^* = \begin{pmatrix} \Lambda_B & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{F}_P.$$

Par l'hypothèse de forte admissibilité primale et duale, et par le théorème 2.4, nous savons que \mathcal{F}_P^* est borné. De plus, la fonction $\log \det X_B$ est strictement concave sur $\text{int}\mathcal{F}_P^*$.

En effet, puisque une fonction est strictement concave si et seulement si toute restriction de cette fonction à une droite est strictement concave, examinons si la fonction $\log \det (X_B + tH)$, où H est une matrice quelconque de même dimension que X_B et où t est un réel, est une fonction de t strictement concave. En fait, nous avons successivement

$$\begin{aligned} \log \det(X_B + tH) &= \log \det(X_B^{\frac{1}{2}}(I + X_B^{-\frac{1}{2}}tHX_B^{-\frac{1}{2}})X_B^{\frac{1}{2}}) \\ &= \log \det X_B + \log \det(I + X_B^{-\frac{1}{2}}tHX_B^{-\frac{1}{2}}) \\ &= \log \det X_B + \sum_{i=1}^k \log(1 + t\lambda_i) \end{aligned}$$

où λ_i représente la $i^{\text{ème}}$ valeur propre de la matrice $X_B^{-\frac{1}{2}}HX_B^{-\frac{1}{2}}$. Or, comme la fonction $\log(1+t\lambda_i)$ est strictement concave en t , la fonction $\log \det(X_B + tH)$ est également strictement concave en t . Ainsi, la fonction $\log \det X_B$ est strictement concave sur $\text{int}\mathcal{F}_P^*$.

Par conséquent, puisque $\log \det X_B$ est strictement concave sur $\text{int}\mathcal{F}_P^*$, ensemble borné, il existe un maximum unique de $\log \det X_B$ sur $\text{int}\mathcal{F}_P^*$. Ce maximum est appelé centre analytique de $\text{int}\mathcal{F}_P^*$ et est noté X^a .

Caractérisons X^a par le système de Karush-Kuhn-Tucker correspondant à ce problème d'optimisation. Nous avons

$$\max \log \det X_B \quad \text{s.c.} \quad X_B \in (\Lambda_B + \mathcal{A}_B) \cap \mathcal{H}_{(k)}^+, \quad X_U = 0 \quad \text{et} \quad X_N = 0. \quad (4.8)$$

Or, nous savons que les conditions de K.K.T. associé à un système tel

$$\max f(x) \quad \text{s.c.} \quad x \in C$$

sont $\nabla f(x^*) \in N(C, x^*)$ où $N(C, x^*)$ est le cône normal à C au point x^* . Nous calculons donc $\nabla f(x^*)$ où $f : \mathcal{H}_{(k)}^{++} \rightarrow R : X_B \rightarrow f(X_B) = \log \det X_B$.

Soient $X_B \in \mathcal{H}_{(k)}^{++}$ et H une matrice carrée quelconque de dimension k . Pour t suffisamment petit, nous avons $X_B + tH \in \mathcal{H}_{(k)}^{++}$. Soient λ_i ($i = 1 \dots k$) les valeurs propres de $X_B^{-\frac{1}{2}} H X_B^{-\frac{1}{2}}$. Par le développement précédent, nous obtenons

$$\frac{f(X_B + tH) - f(X_B)}{t} = \sum_{i=1}^k \frac{\log(1 + t\lambda_i)}{t}.$$

Puisque

$$\lim_{t \rightarrow 0} \frac{\log(1 + t\lambda_i)}{t} = \lim_{t \rightarrow 0} \frac{\lambda_i}{1 + t\lambda_i} = \lambda_i,$$

nous avons $\text{tr} f'(X_B)H = \sum_{i=1}^k \lambda_i = \text{tr} X_B^{-\frac{1}{2}} H X_B^{-\frac{1}{2}} = \text{tr} X_B^{-1} H$. Vu que H est choisie de façon arbitraire, nous en déduisons que $f'(X_B) = X_B^{-1} \in \mathcal{H}_{(k)}^{++}$.

De plus, nous avons $N(\Lambda_B + \mathcal{A}_B, X_B^a) = \mathcal{A}_B^\perp$ puisque $\Lambda_B + \mathcal{A}_B$ est un sous-espace linéaire affine.

Le système de K.K.T. relatif au problème d'optimisation (4.8) est donc

$$\begin{cases} (X_B^a)^{-1} = Z_B & (X_B^a Z_B = I) \\ Z_B \in \mathcal{A}_B^\perp \cap \mathcal{H}_{(k)}^{++} \\ X_B^a \in (\Lambda_B + \mathcal{A}_B) \cap \mathcal{H}_{(k)}^{++} \\ X_U = 0 & X_N = 0. \end{cases} \quad (4.9)$$

Il caractérise le centre analytique X^a de $\mathcal{F}_\mathcal{D}^*$.

De la même manière, nous pouvons définir le centre analytique de $\mathcal{F}_\mathcal{D}^*$. En effet, nous avons

$$\mathcal{A}_N^\perp = \{Z_N \in \mathcal{H}^{(\bar{n}-k)} \mid \begin{pmatrix} 0 & 0 \\ 0 & Z_N \end{pmatrix} \in \mathcal{A}^\perp\},$$

$$\mathcal{F}_\mathcal{D}^* = \{Z \in \mathcal{H}^{(\bar{n})} \mid Z_N \in (\Lambda_N + \mathcal{A}_N^\perp) \cap \mathcal{H}_{(\bar{n}-k)}^+, \quad Z_B = 0 \text{ et } Z_U = 0\}.$$

De nouveau, nous savons par l'hypothèse de forte admissibilité primale et duale et par le théorème 2.4 que $\mathcal{F}_\mathcal{D}^*$ est borné. De plus, la fonction $\log \det Z_N$ est strictement concave sur $\text{int} \mathcal{F}_\mathcal{D}^*$. Par conséquent, il existe un maximum unique de $\log \det Z_N$ sur $\text{int} \mathcal{F}_\mathcal{D}^*$. Celui-ci est appelé le centre analytique de $\mathcal{F}_\mathcal{D}^*$ et est noté Z^a . Le système de K.K.T qui le caractérise est

$$\begin{cases} X_N Z_N^a = I \\ X_N \in \mathcal{A}_N \cap \mathcal{H}_{(\bar{n}-k)}^{++} \\ Z_N^a \in (\Lambda_N + \mathcal{A}_N^\perp) \cap \mathcal{H}_{(\bar{n}-k)}^{++} \\ Z_B = 0 & Z_U = 0 \end{cases} \quad (4.10)$$

Etudions à présent la convergence des μ -centres vers (X^a, Z^a) .

4.2 Convergence des points de la trajectoire centrale primale-duale vers le centre analytique de l'ensemble des solutions optimales primales-duales

Une propriété bien connue en programmation linéaire est la convergence de la trajectoire centrale primale-duale vers le centre analytique de l'ensemble des solutions optimales primales-duales. Or, la programmation linéaire est un cas particulier de la programmation semi-définie. C'est pourquoi, nous aimerions étendre cette caractéristique, prouvée en programmation linéaire, à la programmation semi-définie. Dans cette section, nous effectuons cette preuve. De plus, nous montrons que la distance du centre analytique à tout μ -centre converge vers 0 de la même manière que μ tend vers 0.

Avant de démontrer ces deux résultats, nous admettons deux lemmes pour ne pas alourdir les démonstrations déjà longues.

Lemme 4.1

$$\forall \mu \in]0, 1[\quad \begin{array}{ll} X_B(\mu) = \Theta(1) & X_N(\mu) = \Theta(\mu) \\ Z_B(\mu) = \Theta(\mu) & Z_N(\mu) = \Theta(1) \end{array}$$

et tout point limite de la trajectoire centrale $\{(X(\mu), Z(\mu)) | \mu > 0\}$ lorsque μ tend vers 0 est une paire de solutions primales-duales optimales strictement complémentaires.

Lemme 4.2

$$\forall \mu \in]0, 1[\quad \begin{array}{l} \|X_U(\mu)\| = \Theta(1)\|Z_U(\mu)\| \\ -X_U(\mu) \bullet Z_U(\mu) = \Theta(1)\|X_U(\mu)\|^2 \\ \|X_U(\mu)\| = o(\sqrt{\mu}) \quad \text{lorsque } \mu \rightarrow 0 \\ \|Z_U(\mu)\| = o(\sqrt{\mu}) \quad \text{lorsque } \mu \rightarrow 0. \end{array}$$

Démontrons la convergence des μ -centres vers le couple des centres analytiques (X^a, Z^a) .

Théorème 4.1 La trajectoire centrale primale-duale $\{(X(\mu), Z(\mu)) | \mu > 0\}$ converge vers le centre analytique (X^a, Z^a) de $\text{int}\mathcal{F}_P^* \times \text{int}\mathcal{F}_D^*$.

De plus, si nous posons $\epsilon(\mu) := \frac{\|X_U(\mu)\|}{\sqrt{\mu}}$, alors nous obtenons

$$\|X_B(\mu) - X_B^a\| = O((\epsilon(\mu) + \sqrt{\mu})^2) \quad \text{et} \quad \|Z_N(\mu) - Z_N^a\| = O((\epsilon(\mu) + \sqrt{\mu})^2).$$

preuve : Supposons $0 < \mu < 1$.

La première partie de la démonstration s'intéresse à la convergence des μ -centres vers (X^a, Z^a) . Dans cette optique, nous essayons de trouver un système d'équations, fonction de $X(\mu)$, qui lorsque μ tend vers 0 devient le système caractérisant X^a , (4.9). De même, pour le cas dual, nous désirons trouver un système d'équations, fonction de $Z(\mu)$, qui lorsque μ tend vers 0 devient le système de K.K.T. caractérisant Z^a , (4.10).

Nous nous préoccupons dans un premier temps du cas primal. Par définition de μ -centre, nous avons $X(\mu)Z(\mu) = \mu I$. Si nous partitionnons $X(\mu), Z(\mu)$

et I , cette relation devient

$$\begin{pmatrix} X_B(\mu) & X_U(\mu) \\ X_U(\mu)^H & X_N(\mu) \end{pmatrix} \begin{pmatrix} Z_B(\mu) & Z_U(\mu) \\ Z_U(\mu)^H & Z_N(\mu) \end{pmatrix} = \begin{pmatrix} \mu I_B & \mu I_U \\ \mu I_U^H & \mu I_N \end{pmatrix}.$$

Ceci nous permet d'écrire $X_B(\mu)Z_B(\mu) + X_U(\mu)Z_U(\mu)^H = \mu I_B$. Or, nous avons que $\mu > 0$ et $X_B(\mu) \in \mathcal{H}_k^{++}$. Dès lors, $\mu X_B(\mu)$ est inversible et nous obtenons la relation suivante :

$$X_B(\mu)^{-1} = \frac{1}{\mu} Z_B(\mu) + \frac{1}{\mu} X_B(\mu)^{-1} X_U(\mu) Z_U(\mu)^H. \quad (4.11)$$

De plus, nous savons que

$$\mathcal{A}_B^\perp = \{Z_B \in \mathcal{H}^{(k)} \mid \begin{pmatrix} Z_B & Z_U \\ Z_U^H & Z_N \end{pmatrix} \in \mathcal{A}^\perp\}.$$

Soit $\{\mathcal{A}_B^{(i)} \mid i = 1 \dots m\}$, où $m = \dim \mathcal{A}_B^\perp$, un ensemble de matrices qui engendre le sous-espace linéaire \mathcal{A}_B^\perp . Nous avons que

$$Z_B(\mu) \in \mathcal{A}_B^\perp. \quad (4.12)$$

En effet, nous avons $Z(\mu)$ et Z^* qui appartiennent à $(C + \mathcal{A}^\perp) \cap \mathcal{H}_n^{++}$. Ceci entraîne que $Z(\mu) \in Z^* + \mathcal{A}^\perp$. En partitionnant, nous avons $Z_B(\mu) \in \mathcal{A}_B^\perp$.

Nous obtenons ainsi la relation suivante :

$$\frac{1}{\mu} Z_B(\mu) = \sum_{i=1}^m \nu_i(\mu) \mathcal{A}_B^{(i)} = X_B(\mu)^{-1} - \frac{1}{\mu} X_B(\mu)^{-1} X_U(\mu) Z_U(\mu)^H \quad (4.13)$$

par (4.11) et où $\nu_i(\mu)$ sont des scalaires

De plus, le lemme 4.1 affirme que $\frac{Z_B(\mu)}{\mu} = \Theta(1)$. Par définition de Θ , voir annexe, nous obtenons

$$\begin{aligned} \exists \Gamma > 0 \quad \frac{1}{\Gamma} I &\preceq \frac{Z_B(\mu)}{\mu} \preceq \Gamma I, \quad \mu \in]0, 1[, \\ \frac{1}{\Gamma} I &\preceq \sum_{i=1}^m \nu_i(\mu) \mathcal{A}_B^{(i)} \preceq \Gamma I. \end{aligned}$$

Définissons l'ensemble B comme étant,

$$B = \{(\nu_1(\mu), \dots, \nu_m(\mu)) \in \mathbb{R}^m \mid \frac{1}{\Gamma} I \preceq \sum_{i=1}^m \nu_i(\mu) \mathcal{A}_B^{(i)} \preceq \Gamma I\},$$

et prouvons que celui-ci est borné. De cette manière, nous démontrons l'existence de la limite de $\sum_{i=1}^m \nu_i(\mu) \mathcal{A}_B^{(i)}$ lorsque μ tend vers zéro, indispensable à notre preuve.

Montrons donc que B est inclus dans un ensemble borné. Autrement dit, prouvons que

$$C = \{(l_1, \dots, l_m) \in \mathbb{R}^m \mid \|\sum_{i=1}^m l_i \mathcal{A}_B^{(i)}\|_2 \leq \Gamma\}$$

est borné. Dans ce but, définissons l'application g de domaine de définition $\text{span}(\mathcal{A}_B^{(1)}, \dots, \mathcal{A}_B^{(m)})$ et à valeurs dans R^m qui, à toute matrice de $\text{span}(\mathcal{A}_B^{(1)}, \dots, \mathcal{A}_B^{(m)})$, fait correspondre l'unique vecteur (l_1, \dots, l_m) tel que $A = \sum_{i=1}^m l_i \mathcal{A}_B^{(i)}$. Ceci est rendu possible par le fait que les matrices $\mathcal{A}_B^{(i)}$ $i=1\dots m$ sont linéairement indépendantes. Premièrement, nous avons que l'application g est linéaire et donc continue vu que nous travaillons en dimension finie. Deuxièmement, nous savons que

$$C = g(\{A \mid \|A\|_2 \leq \Gamma\}).$$

Par conséquent, puisque l'image d'un fermé borné par une application continue est fermée bornée, C est borné. Comme relevé ci-dessus, ceci entraîne que B est borné.

Ensuite, les lemmes 4.1 et 4.2, ainsi que la définition de o (voir annexe) entraînent que

$$\lim_{\mu \rightarrow 0} \frac{1}{\mu} X_B(\mu)^{-1} X_U(\mu) Z_U(\mu)^H = 0.$$

En effet, $\|X_U(\mu)\| = \|Z_U(\mu)^H\| = o(\sqrt{\mu})$ lorsque μ tend vers 0 (lemme 4.2), et $X_B(\mu)^{-1} = \Theta(1)$ (lemme 4.1).

Des considérations ci-dessus et de (4.13), nous tirons que lorsque μ tend vers 0, $X_B(\mu)^{-1}$ tend vers $\sum_{i=1}^m \nu_i(\mu) \mathcal{A}_B^{(i)}$.

D'un autre côté, nous savons également que $X(\mu) \in (B + \mathcal{A})$. Soit $\mathcal{A}^{(i)} = \begin{pmatrix} \mathcal{A}_B^{(i)} & \mathcal{A}_U^{(i)} \\ \mathcal{A}_U^{H(i)} & \mathcal{A}_N^{(i)} \end{pmatrix} \in \mathcal{A}^\perp$ pour $i = 1\dots m$. Nous avons successivement

$$\begin{pmatrix} \mathcal{A}_B^{(i)} & \mathcal{A}_U^{(i)} \\ \mathcal{A}_U^{H(i)} & \mathcal{A}_N^{(i)} \end{pmatrix} \bullet \begin{pmatrix} X_B(\mu) & X_U(\mu) \\ X_U(\mu)^H & X_N(\mu) \end{pmatrix} = b_i,$$

$$\mathcal{A}_B^{(i)} \bullet X_B(\mu) + \mathcal{A}_U^{(i)} \bullet X_U(\mu)^H + \mathcal{A}_U^{H(i)} \bullet X_U(\mu) + \mathcal{A}_N^{(i)} \bullet X_N(\mu) = b_i. \quad (4.14)$$

Finalement, de tout ce qui précède, nous pouvons dire que la limite $(X^*, \nu_1^*, \dots, \nu_m^*)$ de $(X(\mu), \nu_1(\mu), \dots, \nu_m(\mu))$ pour μ tendant vers 0 satisfait le système d'équations non linéaire suivant :

$$\begin{cases} X_B^{*-1} - \sum_{i=1}^m \nu_i^* \mathcal{A}_B^{(i)} = 0 \\ \mathcal{A}_B^{(i)} \bullet X_B^* = b_i \quad \text{pour } i = 1\dots m. \end{cases} \quad (4.15)$$

Afin de compléter ce système d'équations, nous utilisons le lemme 4.1 qui affirme que $\frac{Z_B(\mu)}{\mu} = \Theta(1)$ et $X_B(\mu) = \Theta(1)$ où $\mu \in]0, 1[$. Ceci entraîne que

$$\begin{aligned} \exists \Gamma > 0 \quad \frac{1}{\Gamma} I &\preceq \sum_{i=1}^m \nu_i(\mu) \mathcal{A}_B^{(i)} \preceq \Gamma I, \\ \exists \Gamma' > 0 \quad \frac{1}{\Gamma'} I &\preceq X_B(\mu) \preceq \Gamma' I \end{aligned}$$

où $\mu \in]0, 1[$, et donc que

$$\begin{cases} \sum_{i=1}^m \nu_i^* \mathcal{A}_B^{(i)} \succ 0 \\ X_B^* \succ 0. \end{cases} \quad (4.16)$$

Par conséquent, X^* caractérisé par (4.15) et (4.16) est le centre analytique de \mathcal{F}_P^* , puisqu'en posant $Z_B = \lim_{\mu \rightarrow 0} \frac{Z_B(\mu)}{\mu}$, nous retrouvons le système de K.K.T. (4.9). Nous concluons dès lors que $X(\mu) - X^a = o(1)$ pour μ proche de 0.

De façon similaire, nous pouvons conclure que $Z(\mu) - Z^a = o(1)$ pour μ proche de 0.

Dans la seconde partie de la démonstration, nous désirons, après avoir posé $\epsilon(\mu) := \frac{\|X_U(\mu)\|}{\sqrt{\mu}}$, obtenir

$$\|X_B(\mu) - X_B^a\| = O((\epsilon(\mu) + \sqrt{\mu})^2) \quad \text{et} \quad \|Z_N(\mu) - Z_N^a\| = O((\epsilon(\mu) + \sqrt{\mu})^2).$$

Intéressons-nous dans un premier temps à $X_B(\mu)$. Nous voulons appliquer le théorème de la fonction inverse au système linéaire (4.15) au point $X_B = X_B^a$ et $\nu_i = \nu_i^*$ pour $i = 1 \dots m$. Les hypothèses de ce théorème sont satisfaites puisque, par la page 125 de [14], nous avons que le jacobien de ce système en ces points, $J(X_B^a, \nu_1^*, \dots, \nu_m^*)$, est régulier. Il est indiqué également que, par ce théorème, nous obtenons

$$\|X_B(\mu) - X_B^a\| = O(\|X_B(\mu)^{-1} - \sum_{i=1}^m \nu_i(\mu) \mathcal{A}_B^{(i)}\| + \sum_{i=1}^m |\mathcal{A}_B^{(i)} \bullet X_B(\mu) - b_i|).$$

Or, nous avons successivement

$$\begin{aligned} \|X_B(\mu)^{-1} - \sum_{i=1}^m \nu_i(\mu) \mathcal{A}_B^{(i)}\| &= \left\| \frac{1}{\mu} X_B(\mu)^{-1} X_U(\mu) Z_U(\mu)^H \right\| \quad \text{par (4.13)} \\ &= \|X_B(\mu)^{-1} \frac{X_U(\mu)}{\sqrt{\mu}} \frac{Z_U(\mu)^H}{\sqrt{\mu}}\| \\ &\leq \|X_B(\mu)^{-1}\| \left\| \frac{X_U(\mu)}{\sqrt{\mu}} \right\|^2 \quad \text{par le lemme 4.2} \\ &= \|X_B(\mu)^{-1}\| \epsilon(\mu)^2 \\ &= O(\epsilon(\mu)^2). \end{aligned}$$

De plus, l'inégalité suivante est satisfaite par l'intermédiaire de (4.14) :

$$\sum_{i=1}^m |\mathcal{A}_B^{(i)} \bullet X_B(\mu) - b_i| = |2\mathcal{A}_U^{(i)} \bullet X_U(\mu) + \mathcal{A}_N^{(i)} \bullet X_N(\mu)|.$$

Dès lors, nous obtenons successivement

$$\begin{aligned} |\mathcal{A}_B^{(i)} \bullet X_B(\mu) - b_i| &\leq 2|\mathcal{A}_U^{(i)} \bullet X_U(\mu)| + |\mathcal{A}_N^{(i)} \bullet X_N(\mu)| \\ &\leq K\|X_U(\mu)\| + K'\mu \quad \text{par le lemme 4.1} \\ &= o(\epsilon(\mu)\sqrt{\mu} + \mu) \end{aligned}$$

pour $i = 1 \dots m$. Par conséquent, nous concluons que

$$\begin{aligned}\|X_B(\mu) - X_B^a\| &= O(\epsilon(\mu)^2 + \epsilon(\mu)\sqrt{\mu} + \mu) \\ &= O((\epsilon(\mu) + \sqrt{\mu})^2).\end{aligned}$$

De la même manière, on peut montrer que

$$\|Z_N(\mu) - Z_N^a\| = O((\epsilon(\mu) + \sqrt{\mu})^2).$$

cqfd

Le théorème suivant caractérise de façon plus précise la convergence des points $(X(\mu), Z(\mu))$ de la trajectoire centrale vers le centre analytique (X^a, Z^a) de $\mathcal{F}_P^* \times \mathcal{F}_D^*$ lorsque μ tend vers 0. En effet, il prouve que cette convergence est de l'ordre de μ .

Théorème 4.2 *Soit $\mu \in]0, 1[$. Alors, $\|X(\mu) - X^a\| = O(\mu)$ et $\|Z(\mu) - Z^a\| = O(\mu)$.*

preuve : Intéressons-nous au cas primal : $\|X(\mu) - X^a\| = O(\mu)$. En fait, il est suffisant de montrer $\|X_U(\mu)\| = O(\mu)$. En effet, nous écrivons successivement

$$\begin{aligned}\|X_B(\mu) - X_B^a\| &= O((\epsilon(\mu) + \sqrt{\mu})^2) \quad \text{par le théorème 4.1} \\ &= O\left(\left(\frac{\|X_U(\mu)\|}{\sqrt{\mu}} + \sqrt{\mu}\right)^2\right) \\ &= O((2\sqrt{\mu})^2) \quad \text{si } \|X_U(\mu)\| = O(\mu) \\ &= O(\mu) \\ \|X_U(\mu) - X_U^a\| &= \|X_U(\mu)\| = O(\mu) \\ \|X_N(\mu) - X_N^a\| &= \|X_N(\mu)\| = O(\mu)\end{aligned}$$

par le lemme 4.1 et voir annexe O.

Prouvons $\|X_U(\mu)\| = O(\mu)$ par l'absurde. Supposons donc qu'il existe une suite

$$\{(X(\mu_k), Z(\mu_k)) | k = 1, 2, \dots\}$$

avec $\|X_U(\mu_k)\| > 0$ et $\lim_{k \rightarrow \infty} \frac{\mu_k}{\|X_U(\mu_k)\|} = 0$. Nous choisissons la suite $\{(X(\mu_k), Z(\mu_k)) | k = 1, 2, \dots\}$ de telle manière que $\lim_{k \rightarrow \infty} \frac{Z_B(\mu_k)}{\mu_k}$ existe, ce qui est toléré par le lemme 4.1. De plus, définissons

$$\Delta_B^x(\infty) := \lim_{k \rightarrow \infty} \frac{\mu_k}{\|X_U(\mu_k)\|^2} (X_B(\mu_k) - X_B^a), \quad (4.17)$$

ce qui est possible puisque cette limite existe. En effet,

$$\Delta_B^x(\infty) = \lim_{k \rightarrow \infty} \frac{1}{\epsilon(\mu_k)^2} O((\epsilon(\mu_k) + \sqrt{\mu_k})^2)$$

par le théorème 4.1 et par (4.17). Or, nous avons successivement $\frac{\epsilon(\mu_k) + \sqrt{\mu_k}}{\epsilon(\mu_k)} = 1 + \frac{\sqrt{\mu_k}}{\epsilon(\mu_k)} = 1 + \frac{\mu_k}{\|X_U(\mu_k)\|}$, et donc, par hypothèse, nous obtenons $\lim_{k \rightarrow \infty} \frac{\epsilon(\mu_k) + \sqrt{\mu_k}}{\epsilon(\mu_k)} = 1$, c'est-à-dire $\lim_{k \rightarrow \infty} \epsilon(\mu_k) + \sqrt{\mu_k} = \lim_{k \rightarrow \infty} \epsilon(\mu_k)$. Par conséquent, par définition de O , $\Delta_B^x(\infty) = \lim_{k \rightarrow \infty} \frac{1}{\epsilon(\mu_k)^2} \epsilon(\mu_k)^2 K$.

Dès lors, nous obtenons successivement,

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{(X_B(\mu_k) - X_B^a) \bullet Z_B(\mu_k)}{\|X_U(\mu_k)\|^2} &= \lim_{k \rightarrow \infty} \frac{\Delta_B^x(\infty) \bullet Z_B(\mu_k)}{\mu_k} \\ &= \lim_{k \rightarrow \infty} \frac{\Delta_B^x(\infty) \bullet (Z_B(\mu_k) - Z_B^a)}{\mu_k} \quad \text{car } Z_B^a = 0 \\ &= \lim_{j \rightarrow \infty} \frac{1}{\mu_k} \frac{\mu_j}{\|X_U(\mu_j)\|^2} ((X_B(\mu_j) - X_B^a) \bullet (Z_B(\mu_k) - Z_B^a)) \quad \text{par (4.17)} \end{aligned}$$

et donc, nous avons que

$$\lim_{k \rightarrow \infty} \frac{(X_B(\mu_k) - X_B^a) \bullet Z_B(\mu_k)}{\|X_U(\mu_k)\|^2} = 0 \quad (4.18)$$

car $X_B(\mu_j) - X_B^a \in \mathcal{A}_B$ et $Z_B(\mu_k) - Z_B^a \in \mathcal{A}_B^\perp$ par (4.12).

De manière analogue, nous pouvons montrer que

$$\lim_{k \rightarrow \infty} \frac{X_N(\mu_k) \bullet (Z_N(\mu_k) - Z_N^a)}{\|X_U(\mu_k)\|^2} = 0 \quad (4.19)$$

Par conséquent, nous concluons que

$$\begin{aligned} 0 &= \lim_{k \rightarrow \infty} \frac{(X(\mu_k) - X^a) \bullet (Z(\mu_k) - Z^a)}{\|X_U(\mu_k)\|^2} \quad \text{car } X(\mu_k) - X^a \in \mathcal{A} \text{ et } Z(\mu_k) - Z^a \in \mathcal{A}^\perp \\ &= \lim_{k \rightarrow \infty} \frac{1}{\|X_U(\mu_k)\|^2} (2X_U(\mu_k)^H \bullet Z_U(\mu_k) + (X_B(\mu_k) - X_B^a) \bullet Z_B(\mu_k) + X_N(\mu_k) \bullet (Z_N(\mu_k) - Z_N^a)) \end{aligned}$$

car

$$X(\mu_k) - X^a = \begin{pmatrix} X_B(\mu_k) - X_B^a & X_U(\mu_k) \\ X_U(\mu_k)^H & X_N(\mu_k) \end{pmatrix} \quad \text{et} \quad Z(\mu_k) - Z^a = \begin{pmatrix} Z_B(\mu_k) & Z_U(\mu_k) \\ Z_U(\mu_k)^H & Z_N(\mu_k) - Z_N^a \end{pmatrix}$$

ce qui contredit, compte tenu de (4.18) et (4.19), le lemme 4.2.

cqfd

Ainsi, les μ -centres $(X(\mu), Z(\mu))$ tendent vers le centre analytique (X^a, Z^a) de $\mathcal{F}_p^* \times \mathcal{F}_d^*$, lorsque μ tend vers 0, avec un ordre de grandeur μ . Ce résultat est d'une grande importance. En effet, il permet de prouver la convergence vers une solution optimale des algorithmes basés sur une méthode primale-duale de points intérieurs et de suivi de chemin, pour ce qui concerne la résolution d'un problème de programmation semi-définie.

4.3 Analyse de la convergence d'un algorithme de type prédicteur-correcteur

Nous proposons dans cette section un algorithme primal-dual de points intérieurs et de suivi de chemin de type prédicteur-correcteur. Comme nous l'avons réalisé pour les algorithmes détaillés dans le chapitre 3, nous analysons la centralité des itérés, leur saut de dualité et la complexité algorithmique. Nous terminons par la preuve de la convergence superlinéaire d'ordre $\frac{2}{1+2^{-r}}$, où r est un paramètre, de $(X^{(k)}, Z^{(k)})$ vers le centre analytique (X^a, Z^a) de $\mathcal{F}_P^* \times \mathcal{F}_D^*$ ainsi que la convergence superlinéaire de μ_k vers 0 avec le même ordre.

4.3.1 Algorithme SDP(ϵ) de type prédicteur-correcteur

L'algorithme SDP(ϵ) est un algorithme qui suit la démarche générale d'une méthode primale-duale de points intérieurs et de suivi de chemin pour la résolution d'un problème de programmation semi-définie. Celle-ci est exposée dans la section 3.5.1. Les résultats généraux des chapitres précédents restent donc toujours d'application. Cet algorithme prédicteur-correcteur possède beaucoup de similitudes avec l'algorithme de ce même type détaillé en 3.5.3. Dans cette section, nous analysons cet algorithme et calculons sa complexité algorithmique.

L'algorithme SDP(ϵ) se caractérise de la manière suivante. Nous choisissons $\mathcal{N}(\cdot) = \mathcal{N}_2(\frac{1}{2})$ comme précédemment pour 3.5.3. Nous prenons également $\gamma^{(k)} = 0$ lorsque s'effectue un pas prédicteur et $\gamma^{(k)} = 1$ lorsque se réalise un pas correcteur. Nous considérons toujours une alternance de pas prédicteurs et de pas correcteurs. Néanmoins, nous introduisons un paramètre r correspondant à une borne supérieure du nombre de pas correcteurs successifs entre deux pas prédicteurs. Il n'y a donc plus nécessairement un seul pas correcteur entre deux pas prédicteurs.

L'algorithme est le suivant :

Etant donné $(X^{(0)}, Z^{(0)}) \in \mathcal{F}_P \times \mathcal{F}_D$ tel que $\delta(V^{(0)}) \leq \frac{1}{4}$, ϵ tel que $0 < \epsilon \leq \frac{X^{(0)} \bullet Z^{(0)}}{\bar{n}}$ et r un entier positif ≥ 1 ,

PAS 0 $k := 0$

PAS 1 si $X^{(k)} \bullet Z^{(k)} < \epsilon$ alors l'algorithme se termine puisque le saut de dualité est assez petit.

PAS 2 $X := X^{(k)}$, $Z := Z^{(k)}$ et $\mu^{(k)} := \frac{X \bullet Z}{\bar{n}}$

– Pas prédicteur :

$\gamma = 0$

calculer $(\Delta X^p, \Delta Z^p)$ en résolvant le système linéaire (3.20)

$t^{(k)} = \max\{\bar{t} \mid \delta(V(t)) \leq \min(\frac{1}{2}, ((1-t)\frac{\mu^{(k)}}{\epsilon})^{2^{-r}}) \quad \forall t \in [0, \bar{t}]\}$

$$X' := X + t^{(k)} \Delta X^p, \quad Z' := Z + t^{(k)} \Delta Z^p \text{ et } \beta^{(k)} = \min\left(\frac{1}{4}, (1 - t^{(k)}) \frac{\mu^{(k)}}{\epsilon}\right)$$

– Pas correcteur(s) :

pour $i=1$ jusque r :

$\gamma=1$, $X := X'$, $Z := Z'$, calculer V

si $\delta(V) \leq \beta^{(k)}$ alors aller au pas 3

calculer $(\Delta X^c, \Delta Z^c)$ en résolvant le système linéaire (3.20)

$X' := X + \Delta^c$ et $Z' := Z + \Delta^c$

PAS 3 $X^{(k+1)} = X'$, $Z^{(k+1)} = Z'$ et $k := k + 1$

retour au pas 1.

Commentons quelque peu cet algorithme et comparons-le à l'algorithme prédicteur-correcteur de 3.5.3. La différence majeure est, nous le savons, que l'algorithme 3.5.3. comporte un seul pas correcteur entre deux pas prédicteurs tandis que $\text{SDP}(\epsilon)$ peut comprendre jusqu'à r pas correcteurs entre deux pas prédicteurs.

Examinons ce qu'il advient de la centralité des itérés et du saut de dualité. Considérons le début de l'itération k ($k \geq 1$). Les itérés sont $(X^{(k)}, Z^{(k)})$ avec un saut de dualité de $\bar{n}\mu^{(k)}$.

En premier lieu, nous effectuons un pas prédicteur et obtenons les points (X', Z') . Du point de vue de la centralité nous voyons que la définition de la longueur $t^{(k)}$ de ce pas entraîne la propriété suivante : $\delta(V'(t^{(k)})) \leq \min(\frac{1}{2}, ((1 - t^{(k)}) \frac{\mu^{(k)}}{\epsilon})^{2-r})$. Nous concluons de façon identique à l'algorithme 3.5.3 à l'appartenance de $V'(t^{(k)})$ au voisinage $\mathcal{N}_2(\frac{1}{2})$. De plus, par le même raisonnement que celui développé pour l'algorithme 3.5.3, nous prouvons que la longueur du pas prédicteur est supérieure ou égale à $\frac{2}{1+\sqrt{1+8\bar{n}}}$. Néanmoins, nous avons une contrainte supplémentaire pour $\text{SDP}(\epsilon)$, $\delta(V'(t^{(k)}))$ doit être inférieur à $((1 - t^{(k)}) \frac{\mu^{(k)}}{\epsilon})^{2-r}$. L'utilité de celle-ci sera développée plus bas. L'intérêt de ce pas ne réside donc pas en une meilleure centralité des itérés.

Par contre du point de vue du saut de dualité, nous avons successivement

$$\begin{aligned} X' \bullet Z' &= \bar{n}(1 - t^{(k)})\mu^{(k)} \\ &= (1 - t^{(k)})X^{(k)} \bullet Z^{(k)}. \end{aligned}$$

Le saut de dualité diminue ainsi strictement puisque $t^{(k)} \geq \frac{2}{1+\sqrt{1+8\bar{n}}} > 0$. Le pas prédicteur vise une solution optimale et ne se préoccupe nullement de la proximité de la trajectoire centrale.

En second lieu, nous effectuons de 0 à r pas correcteurs à partir du point (X', Z') afin d'obtenir les itérés suivants $(X^{(k+1)}, Z^{(k+1)})$. Envisageons le point de vue du saut de dualité, nous avons

$$X^{(k+1)} \bullet Z^{(k+1)} = \bar{n}\mu' = X' \bullet Z'$$

où $\mu' = (1 - t^{(k)})\mu^{(k)}$, puisque de façon identique à l'algorithme 3.5.3, le pas correcteur conserve le saut de dualité.

Par contre du point de vue de la centralité. Nous voyons que l'algorithme $\text{SDP}(\epsilon)$ réalise autant de pas correcteurs nécessaires à l'obtention de la propriété suivante :

$$\delta(V^{(k+1)}) \leq \beta^{(k)} = \min\left(\frac{1}{4}, (1 - t^{(k)}) \frac{\mu^{(k)}}{\epsilon}\right) \quad (4.20)$$

où $t^{(k)}$ est la longueur du pas prédicteur. Nous concluons de façon identique à l'algorithme 3.5.3., à l'appartenance de $V^{(k+1)}$ à $\mathcal{N}_2(\frac{1}{4})$. Néanmoins, nous avons pour $\text{SDP}(\epsilon)$ une contrainte supplémentaire,

$$\delta(V^{(k+1)}) \leq (1 - t^{(k)}) \frac{\mu^{(k)}}{\epsilon} = \frac{\mu^{(k+1)}}{\epsilon} = O(\mu^{(k+1)}). \quad (4.21)$$

Notons que (4.20) est obtenue en au plus r pas correcteurs. En effet, le lemme 3.5 affirme que si $\gamma=1$ et $\delta(V) \leq \frac{1}{2}$, alors $\delta(V(1)) \leq \delta(V)^2$. Dans notre cas, nous avons $\gamma=1$ et $t=1$, caractéristiques du pas correcteur. De plus, nous avons $\delta(V) \leq \frac{1}{2}$ puisque nous venons de réaliser soit un pas prédicteur, soit un pas correcteur. Appliquant un premier pas correcteur, nous obtenons de nouveaux points $(X^{(k+1)}, Z^{(k+1)})$, et nous avons par le lemme 3.5.

$$\delta(V^{(k+1)}) \leq \delta(V')^2 \leq \min\left(\frac{1}{4}, ((1 - t^{(k)}) \frac{\mu^{(k)}}{\epsilon})^{2^{-r+1}}\right)$$

Si (4.20) n'est pas satisfaite, nous appliquons un second pas correcteur pour obtenir de nouveaux points, que nous notons également $(X^{(k+1)}, Z^{(k+1)})$. Après r pas correcteurs, se référant au lemme 3.5., nous avons

$$\delta(V^{(k+1)}) \leq \min\left(\frac{1}{2^r}, ((1 - t^{(k)}) \frac{\mu^{(k)}}{\epsilon})\right) \quad (4.22)$$

L'intérêt de ce pas réside visiblement en une meilleure centralité des itérés.

A ce moment, l'itération k est finie. Cette itération nous a permis par le pas prédicteur de réduire le saut de dualité et donc de se rapprocher d'une solution optimale, $X^{(k+1)} \bullet Z^{(k+1)} = (1 - t^{(k)}) X^{(k)} \bullet Z^{(k)}$, et de veiller à ce que $(X^{(k+1)}, Z^{(k+1)})$ soit proche de la trajectoire centrale via le ou les pas correcteurs.

Terminons par l'étude de la complexité algorithmique de $\text{SDP}(\epsilon)$.

Théorème 4.3 $\forall \epsilon$ tel que $0 < \epsilon < \frac{X^{(0)} \bullet Z^{(0)}}{\bar{n}}$, l'algorithme $\text{SDP}(\epsilon)$ calcule une solution α -optimale où $\alpha = \bar{n} \frac{\epsilon}{4}$ en au plus $O(\sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$ itérations.

preuve : Si nous supposons qu'il existe k tel que $\frac{\mu^{(k+1)}}{\epsilon} \geq \frac{1}{4}$, alors, puisque $\beta^{(k)} = \min(\frac{1}{4}, ((1 - t^{(k)}) \frac{\mu^{(k)}}{\epsilon}))$, nous avons $\beta^{(k)} = \frac{1}{4}$. Dans ce cas, un seul pas correcteur est nécessaire. En effet, l'analyse de la centralité pour un pas correcteur nous dit que si $V^{(k)} \in \mathcal{N}_2(\frac{1}{2})$, alors $V^{(k+1)} \in \mathcal{N}_2(\frac{1}{4})$. Par conséquent, puisque le pas prédicteur a permis à $V^{(k)}$ d'appartenir à $\mathcal{N}_2(\frac{1}{2})$ et puisque $\beta^{(k)} = \frac{1}{4}$, un seul pas correcteur est nécessaire à l'obtention de la propriété $\delta(V^{(k+1)}) \leq \beta^{(k)}$ qui détermine le nombre de pas correcteurs.

Dès lors, toutes les itérations k telles que $\frac{\mu^{(k+1)}}{\epsilon} \geq \frac{1}{4}$ de l'algorithme $\text{SDP}(\epsilon)$ sont identiques aux itérations de l'algorithme 3.5.3.

Or, le théorème 3.2 analysant la complexité algorithmique de 3.5.3 prouve que cet algorithme calcule une solution ϵ -optimale en $O(\sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$ itérations. Par conséquent, nous concluons que pour l'algorithme $\text{SDP}(\epsilon)$, nous avons

$$\mu^{(k)} = \frac{X^{(k)} \bullet Z^{(k)}}{\bar{n}} \leq \frac{\epsilon}{4} \quad \forall k \geq \Gamma \sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon}$$

où Γ est une constante. En effet, raisonnons par l'absurde. Supposons qu'il existe $k' - 1 \geq \Gamma \sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon}$ tel que $\mu^{(k')} > \frac{\epsilon}{4}$. L'itération $k' - 1$ de $\text{SDP}(\epsilon)$ est donc identique à l'itération $k' - 1$ de l'algorithme 3.5.3. Or, par le théorème 3.2, puisque $k' - 1 \geq \Gamma \sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon}$, $(X^{(k')}, Z^{(k')})$ est une solution ϵ -optimale, ce qui contredit le fait que $\mu^{(k')} = \frac{X^{(k')} \bullet Z^{(k')}}{\bar{n}} > \frac{\epsilon}{4}$.

Dès lors, l'algorithme $\text{SDP}(\epsilon)$ calcule une solution α -optimale où $\alpha = \frac{\bar{n}\epsilon}{4}$ en $O(\sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$ itérations.

cqfd

4.3.2 Convergence globale de l'algorithme $\text{SDP}(\epsilon)$

Nous savons, tout comme pour l'algorithme prédicteur-correcteur du chapitre 3, que pour $\text{SDP}(\epsilon)$ le pas prédicteur diminue strictement le saut de dualité. En effet, nous avons que $\mu' = (1 - t^{(k)})\mu^{(k)}$ car $\gamma = 0$ et que la longueur de pas $t^{(k)} \geq \frac{2}{1 + \sqrt{1 + 8\bar{n}}}$. Le pas correcteur, par contre, conserve le saut de dualité. En effet, $\mu^{(k+1)} = \mu'$. Par conséquent, à la fin de l'itération k , nous avons $\mu^{(k+1)} = (1 - t^{(k)})\mu^{(k)}$. De ce fait, la suite $\mu^{(0)}, \mu^{(1)}, \dots$ est une suite monotone strictement décroissante.

De plus, puisque $\mu^{(k)}$ est toujours positif, $\lim_{k \rightarrow \infty} \mu^{(k)}$ existe. Le théorème suivant prouve que celle-ci est nulle. Ainsi, les itérés de cet algorithme convergent vers une solution optimale.

Théorème 4.4 *L'algorithme $\text{SDP}(\epsilon)$ converge globalement, c'est-à-dire que $\lim_{k \rightarrow \infty} \mu^{(k)} = 0$.*

preuve : Effectuons une preuve par l'absurde. Supposons donc que $\mu^{(\infty)} = \lim_{k \rightarrow \infty} \mu^{(k)} > 0$.

Afin d'amener une contradiction, nous allons montrer tout d'abord que, pour un certain k , $t^{(k)} \geq \frac{1}{2} \sqrt{\frac{\mu^{(k)}}{\bar{n}\epsilon}}$ où $t^{(k)}$ est la longueur de pas associée à l'étape prédicteur dans l'itération k . Rappelons que $t^{(k)}$ est le plus grand \bar{t} tel que $\forall 0 \leq t \leq \bar{t} \quad \delta(V^{(k)})(t) \leq \min(\frac{1}{2}, ((1 - t)\frac{\mu^{(k)}}{\epsilon})^{2-r})$. Ainsi, nous allons montrer que pour t tel que $0 \leq t \leq \frac{1}{2} \sqrt{\frac{\mu^{(k)}}{\bar{n}\epsilon}}$, nous

avons $\delta(V^{(k)})(t) \leq \min(\frac{1}{2}, ((1-t)\frac{\mu^{(k)}}{\epsilon})^{2^{-r}})$.

Pour cela, considérons une itération k où $k \geq \Gamma\sqrt{\bar{n}}\log\frac{X^{(0)} \bullet Z^{(0)}}{\epsilon}$ et où $\Gamma \geq 1$ et prouvons tout d'abord que pour t tel que $0 \leq t \leq \frac{1}{2}\sqrt{\frac{\mu^{(k)}}{\bar{n}\epsilon}}$, nous avons $\delta(V^{(k)})(t) \leq ((1-t)\frac{\mu^{(k)}}{\epsilon})^{2^{-r}}$. Par le théorème 4.3, nous avons

$$\frac{X^{(k)} \bullet Z^{(k)}}{\bar{n}} = \mu^{(k)} \leq \frac{\epsilon}{4}. \quad (4.23)$$

Or, par (4.20), nous avons successivement

$$\begin{aligned} \delta(V^{(k)}) &\leq \beta^{(k-1)} \\ &= \min(\frac{1}{4}, (1-t^{(k-1)})\frac{\mu^{(k-1)}}{\epsilon}) \quad \text{par SDP}(\epsilon) \\ &= \min(\frac{1}{4}, \frac{\mu^{(k)}}{\epsilon}) \\ &= \frac{\mu^{(k)}}{\epsilon} \quad \text{par (4.23)}. \end{aligned} \quad (4.24)$$

Par ailleurs, si nous considérons une longueur de pas t tel que

$$0 \leq t \leq \frac{1}{2}\sqrt{\frac{\mu^{(k)}}{\bar{n}\epsilon}}, \quad (4.25)$$

grâce au fait que $\frac{\mu^{(k)}}{\epsilon} \leq \frac{1}{4}$, nous obtenons que

$$1-t \geq \frac{3}{4}. \quad (4.26)$$

De plus, nous souvenant du lemme 3.7 et tenant compte de (4.24) et (4.25), nous trouvons

$$\mu(t)\delta(V)(t) \leq (1-t)\mu\delta(V) + t^2\|\mathcal{P}_{\mathcal{H}}(D_X^p D_Z^p)\|_F.$$

Or, puisque nous considérons dans l'itération seulement l'étape prédicteur, le saut de dualité est défini par $\mu(t) = (1-t)\mu$. De plus, il est évident que $\|\mathcal{P}_{\mathcal{H}}(D_X^p D_Z^p)\|_F = \|\frac{D_X^p D_Z^p + (D_X^p D_Z^p)^H}{2}\|_F \leq \|D_X^p D_Z^p\|_F$. Par conséquent, nous avons

$$\delta(V)(t) \leq \delta(V) + \frac{t^2}{1-t} \frac{\|D_X^p D_Z^p\|_F}{\mu}. \quad (4.27)$$

Et donc, il suit que $\delta(V)(t) \leq \delta(V) + \frac{\bar{n}t^2}{2(1-t)}$. En effet, nous avons successivement

$$\begin{aligned} (\|D_X^p\|_F - \|D_Z^p\|_F)^2 &\geq 0 \\ \|D_X^p\|_F^2 + \|D_Z^p\|_F^2 &\geq 2\|D_X^p\|_F\|D_Z^p\|_F \\ &\geq 2\|D_X^p D_Z^p\|_F \quad \text{voir annexes.} \end{aligned}$$

Or, par (3.13), nous avons $\|D_X^p\|_F^2 + \|D_Z^p\|_F^2 = \bar{n}\mu$, donc tenant compte de l'inégalité ci-dessus, nous obtenons $\frac{\|D_X^p D_Z^p\|_F}{\mu} \leq \frac{\bar{n}}{2}$.

Par conséquent, nous écrivons successivement

$$\begin{aligned} \delta(V(t)) &\leq \frac{\mu^{(k)}}{\epsilon} + \frac{\bar{n}t^2}{2(1-t)} \quad \text{par (4.24)} \\ &\leq \frac{\mu^{(k)}}{\epsilon} + \frac{\bar{n}t^2}{\frac{3}{2}} \quad \text{par (4.26)} \\ &\leq \frac{\mu^{(k)}}{\epsilon} + \frac{1}{6} \frac{\mu^{(k)}}{\epsilon} = \frac{7}{6} \frac{\mu^{(k)}}{\epsilon} \quad \text{par (4.25)} \\ &\leq \frac{7}{6} \frac{4}{3} (1-t) \frac{\mu^{(k)}}{\epsilon} \quad \text{par (4.26)} \\ &< \left(\frac{(1-t)\mu^{(k)}}{\epsilon} \right)^{2-r}. \end{aligned}$$

La dernière inégalité est obtenue de la manière suivante. Posons $a = \frac{(1-t)\mu^{(k)}}{\epsilon}$. Puisque $0 < 1-t < 1$ et $0 < \frac{\mu^{(k)}}{\epsilon} \leq \frac{1}{4}$, nous avons $0 < a < \frac{1}{4}$. Donc,

$$\frac{7}{6} \frac{4}{3} a^{(1-\frac{1}{2^r})} \leq \frac{7}{6} \frac{4}{3} a^{\frac{1}{2}}$$

puisque $r \geq 1$, a^x est une fonction décroissante et $(1 - \frac{1}{2^r}) \geq \frac{1}{2} \forall r \geq 1$. Il suit que

$$\frac{7}{6} \frac{4}{3} a^{(1-\frac{1}{2^r})} \leq \frac{7}{6} \frac{4}{3} \frac{1}{2} < 1.$$

D'autre part, pour $0 \leq t \leq \frac{1}{2} \sqrt{\frac{\mu^{(k)}}{\epsilon}}$, nous avons $\delta(V^{(k)})(t) \leq \frac{7}{6} \frac{\mu^{(k)}}{\epsilon} \leq \frac{7}{6} \frac{1}{4} \leq \frac{1}{2}$.

Par conséquent, comme $\text{SDP}(\epsilon)$ choisit le pas prédicteur $t^{(k)}$ comme étant le plus grand \bar{t} satisfaisant à $\forall 0 \leq t \leq \bar{t} \quad \delta(V^{(k)})(t) \leq \min(\frac{1}{2}, (1-t)\frac{\mu^{(k)}}{\epsilon})^{2-r}$, nous avons $t^{(k)} \geq \frac{1}{2} \sqrt{\frac{\mu^{(k)}}{\epsilon}}$.

Dès lors, nous obtenons $t^{(k)} \geq \frac{1}{2} \sqrt{\frac{\mu^{(\infty)}}{\epsilon}} = \Theta(1)$. Cette relation entraîne que $1 - \frac{\mu^{(k+1)}}{\mu^{(k)}} = t^{(k)} \geq \Theta(1)$, c'est-à-dire

$$\exists M > 0 \text{ tel que } \frac{1}{M} \leq 1 - \frac{\mu^{(k+1)}}{\mu^{(k)}} \leq M.$$

Ceci est également valable pour $\mu^{(\infty)}$ qui est strictement positif. Nous obtenons

$$\exists M > 0 \text{ tel que } \frac{1}{M} \leq 0 \leq M,$$

qui est impossible.

Par conséquent, $\mu^{(\infty)}$ doit être nul.

cqfd

4.3.3 Convergence superlinéaire du saut de dualité vers 0 ainsi que de $(X^{(k)}, Z^{(k)})$ vers le centre analytique (X^a, Z^a)

Nous venons de démontrer la convergence globale de $\text{SDP}(\epsilon)$. Dans cette section, nous allons plus loin : nous établissons la convergence superlinéaire de $\text{SDP}(\epsilon)$. Celle-ci se trouve à deux niveaux : d'une part, la convergence superlinéaire du saut de dualité vers 0 et d'autre part, la convergence superlinéaire de $(X^{(k)}, Z^{(k)})$ vers le centre analytique (X^a, Z^a) de l'ensemble des solutions optimales primales-duales. Nous prouvons également que l'ordre de ces deux convergences est identique et fonction du paramètre r , qui indique la borne supérieure de pas correcteurs possibles entre deux pas prédicteurs.

La preuve du théorème étant fastidieuse, il est nécessaire de la simplifier quelque peu par l'introduction de notations et de lemmes.

Tout d'abord, nous désirons nous aider du théorème 4.2 qui explicite la convergence des μ -centres vers le centre analytique. Considérons donc $\forall \mu \in R^{++}$, le couple $(X(\mu), Z(\mu))$.

Ensuite, comme nous avons recours au V-espace, rappelons que la matrice de transformation associée à une paire de trajectoire $(X(t), Z(t))$, $L_d G(t)$ où $G(t)$ est définie comme

$$G(t) = D(L_d^{-1} X(t) L_d^{-H}, L_d^H Z(t) L_d)^{\frac{1}{2}}$$

permet de transformer nos variables initiales $(X(t), Z(t))$ en une seule variable diagonale avec ses éléments positifs,

$$\begin{aligned} V(t) &= G(t)^{-1} L_d^{-1} X(t) L_d^{-H} G(t)^{-H} = G(t)^H L_d^H Z(t) L_d G(t) \\ &= G(t)^{-1} \bar{X}(t) G(t)^{-H} = G(t)^H \bar{Z}(t) G(t) \end{aligned}$$

où $V^2(t) = G(t)^{-1} L_d^{-1} X(t) Z(t) L_d G(t)$.

Définissons la matrice de transformation associée aux μ -centres $(X(\mu), Z(\mu))$,

$$L_\mu := L_d G(t) \text{ où } G(t) = D(L_d^{-1} X(\mu) L_d^{-H}, L_d^H Z(\mu) L_d)^{\frac{1}{2}}.$$

Or, par définition de μ -centre, $X(\mu)Z(\mu) = \mu I$. Dès lors,

$$L_\mu^{-1} X(\mu) Z(\mu) L_\mu = \mu I \text{ et } L_\mu^{-1} X(\mu) L_\mu^{-H} = \sqrt{\mu} I = L_\mu^H Z(\mu) L_\mu$$

Suivent ci-dessous quatre lemmes nécessaires à la démonstration du théorème. Ceux-ci nous fournissent des ordres de grandeurs.

Lemme 4.3 Si $\delta(V) \leq \frac{1}{2}$, alors $\|(L_d^{-1}(X(\mu) - X) L_d^{-H})\| + \|L_d^H(Z(\mu) - Z) L_d\| = O(\sqrt{\mu} \delta(V))$.

Lemme 4.4 Si $\delta(V) \leq \frac{1}{2}$ et $\bar{D}_\mu = D(L_d^{-1} X(\mu) L_d^{-H}, L_d^H Z(\mu) L_d)$, alors $\|\bar{D}_\mu - I\| = O(\delta(V))$ et $\bar{D}_\mu = \Theta(1)$.

Lemme 4.5 $\|D_X^p D_Z^p\| = O(\mu(\mu + \delta(V)))$.

Lemme 4.6 $\forall \mu > 0 \quad \|X(\mu)\| + \|Z(\mu)\| = O(1 + \mu)$.

Les preuves de ces quatre lemmes se trouvent respectivement aux pages 135, 136, 138, 120 de [14].

Théorème 4.5 (Convergence superlinéaire) *Les itérés $(X^{(k)}, Z^{(k)})$ générés par l'algorithme $SDP(\epsilon)$ convergent vers le centre analytique (X^a, Z^a) superlinéairement à l'ordre $\frac{2}{1+2-r}$. Le saut de dualité $\mu^{(k)}$ converge vers 0 au même taux de convergence.*

preuve : Intéressons-nous tout d'abord au saut de dualité. Puisque le pas correcteur ne modifie pas le saut de dualité, nous avons par le pas prédicteur $\mu^{(k+1)} = (1 - t^{(k)})\mu^{(k)}$. Nous désirons prouver la convergence superlinéaire de $\mu^{(k)}$ avec une vitesse de l'ordre de $\frac{2}{1+2-r}$. Ceci revient donc à démontrer

$$\exists c_k \rightarrow 0 \text{ et } k_0 \in \mathbb{N} \text{ tel que } \forall k \geq k_0, \mu^{(k+1)} \leq c_k (\mu^{(k)})^{\frac{2}{1+2-r}}.$$

Puisque $\mu^{(k+1)} = (1 - t^{(k)})\mu^{(k)}$, c'est en bornant proprement $(1 - t^{(k)})$ que nous parviendrons à démontrer cette convergence.

De (4.27) et $t < 1$, nous avons

$$\delta(V^{(k)})(t) \leq \delta(V^{(k-1)}) + \frac{1}{1-t} \frac{\|(D_X^p D_Z^p)\|_F}{\mu^{(k)}}. \quad (4.28)$$

Considérons tous les t positifs tels que

$$\beta^{(k-1)} + \frac{\|D_X^p D_Z^p\|_F}{\mu^{(k)}} \leq (1-t) \left((1-t) \frac{\mu^{(k)}}{\epsilon} \right)^{2-r}. \quad (4.29)$$

Nous obtenons successivement

$$\begin{aligned} \delta(V^{(k)})(t) &\leq \beta^{(k-1)} + \frac{1}{1-t} \frac{\|(D_X^p D_Z^p)\|_F}{\mu^{(k)}} \text{ par (4.20) et (4.28)} \\ &\leq \beta^{(k-1)} \left(1 - \frac{1}{1-t}\right) + \left((1-t) \frac{\mu^{(k)}}{\epsilon} \right)^{2-r} \text{ par (4.29)} \\ &\leq \left((1-t) \frac{\mu^{(k)}}{\epsilon} \right)^{2-r}. \end{aligned} \quad (4.30)$$

Posons \bar{t} comme étant le plus grand t vérifiant (4.29). En d'autres mots, \bar{t} est tel que

$$\beta^{(k-1)} + \frac{\|D_X^p D_Z^p\|_F}{\mu^{(k)}} = (1-\bar{t}) \left((1-\bar{t}) \frac{\mu^{(k)}}{\epsilon} \right)^{2-r}.$$

Or, par définition de $t^{(k)}$ dans l'étape prédicteur de l'itération k , $t^{(k)}$ est le plus grand satisfaisant

$$\forall t \quad 0 \leq t \leq t^{(k)} \quad \delta(V^{(k)})(t) \leq \min\left(\frac{1}{2}, \left((1-t) \frac{\mu^{(k)}}{\epsilon} \right)^{2-r}\right).$$

Par conséquent, nous avons $\delta(V^{(k)}(t^{(k)})) \leq ((1 - t^{(k)}) \frac{\mu^{(k)}}{\epsilon})^{2-r}$.

Par ailleurs, par (4.30), \bar{t} est aussi tel que $\delta(V^{(k)}(\bar{t})) \leq ((1 - \bar{t}) \frac{\mu^{(k)}}{\epsilon})^{2-r}$. Donc, $t^{(k)} \geq \bar{t}$ et

$$\begin{aligned} (1 - t^{(k)})^{1+2-r} &\leq (1 - \bar{t})^{1+2-r} \\ &= (\beta^{(k-1)} + \frac{\|D_X^p D_Z^p\|_F}{\mu^{(k)}}) (\frac{\mu^{(k)}}{\epsilon})^{-2-r} \text{ par (4.29)} \\ &= O((\mu^{(k)})^{1-2-r}) \end{aligned}$$

puisque

$$\begin{aligned} \beta^{(k-1)} &\leq (1 - t^{(k-1)}) \frac{\mu^{(k-1)}}{\epsilon} = \frac{\mu^{(k)}}{\epsilon} \text{ par (4.21)} \\ &= O(\mu^{(k)}) \\ \text{et } \|D_X^p D_Z^p\|_F &= O(\mu^{(k)}(\mu^{(k)} + \delta(V))) \end{aligned}$$

par le lemme 4.5. Nous concluons donc que

$$\begin{aligned} \mu^{(k+1)} &= (1 - t^{(k)}) \mu^{(k)} = O((\mu^{(k)})^{\frac{1-2-r}{1+2-r}} \mu^{(k)}) \\ &= O(\mu^{(k) \frac{2}{1+2-r}}). \end{aligned}$$

Utilisant la définition de O , nous obtenons la convergence superlinéaire d'ordre $\frac{2}{1+2-r}$ du saut de dualité.

Dans un second temps, préoccupons-nous de la convergence des itérés $(X^{(k)}, Z^{(k)})$ vers le centre analytique (X^a, Z^a) de l'ensemble des solutions optimales primales-duales. Nous voulons obtenir la convergence superlinéaire d'ordre identique à celui du saut de dualité. En d'autres termes, nous désirons obtenir $\|X^{(k)} - X^a\|_F = O(\mu^{(k)})$ pour le cas primal et son analogue dual. Nous n'effectuons la preuve que pour le cas primal, celle pour le dual étant tout à fait similaire.

Dans ce cas, puisque le théorème (4.2) affirme que $\|X(\mu^{(k)}) - X^a\|_F = O(\mu^{(k)})$, nous prouvons que $\|X^{(k)} - X(\mu^{(k)})\|_F = O(\mu^{(k)})$. La conclusion est alors immédiate.

$$\|X^{(k)} - X(\mu^{(k)})\|_F = \text{Re } \text{tr}(X^{(k)} - X(\mu^{(k)})) L_d^{-H} (L_d^H L_d) L_d^{-1} (X^{(k)} - X(\mu^{(k)})).$$

Or, $\text{Re } \text{tr} XY = \text{Re } \text{tr} YX$ par (2.2), donc, nous avons

$$\|X^{(k)} - X(\mu^{(k)})\|_F = \text{Re } \text{tr} \underbrace{(L_d^H L_d)}_A \underbrace{((X^{(k)} - X(\mu^{(k)})) L_d^{-H} L_d^{-1} (X^{(k)} - X(\mu^{(k)})))}_B$$

Or, de [15], nous avons le théorème suivant :

Si A et B sont des matrices complexes $n \times n$, U et V des matrices unitaires. Si $\sigma_1(A) \geq \sigma_2(A) \geq \dots \geq \sigma_n(A) \geq 0$ et $\sigma_1(B) \geq \sigma_2(B) \geq \dots \geq \sigma_n(B) \geq 0$ alors $\text{Re } \text{tr} U A V B \leq |\text{tr} U A V B| \leq \sum_{i=1}^n \sigma_i(A) \sigma_i(B)$.

Dans notre cas, nous posons $U=V=I$. A et B étant hermitienne, nous avons $|\lambda_i(A)| = \sigma_i(A)$ et $|\lambda_i(B)| = \sigma_i(B)$. A et B étant semi-définie positive, nous avons $\lambda_i(A) \geq 0$ et $\lambda_i(B) \geq 0$. Par conséquent, $\lambda_i(A) = \sigma_i(A)$ et $\lambda_i(B) = \sigma_i(B)$. Nous obtenons, par le théorème

$$\begin{aligned} \operatorname{Re} \operatorname{tr} AB &\leq \sum_{i=1}^n \lambda_i(A) \lambda_i(B) \\ &\leq \sum_{i=1}^n \|A\|_2 \lambda_i(B) \\ &= \|A\|_2 \operatorname{tr} B \text{ par (2.2).} \end{aligned}$$

c'est-à-dire que

$$\begin{aligned} \|X^{(k)} - X(\mu^{(k)})\|_F &\leq \|(L_d^H L_d)\|_2 \operatorname{tr}((X^{(k)} - X(\mu^{(k)})) L_d^{-H} L_d^{-1} (X^{(k)} - X(\mu^{(k)}))) \\ &\quad \text{or } \operatorname{tr} AB = \operatorname{tr} BA \\ &= \|(L_d^H L_d)\|_2 \operatorname{tr}(L_d^{-1} (X^{(k)} - X(\mu^{(k)})) L_d^{-H} (L_d^H L_d) L_d^{-1} (X^{(k)} - X(\mu^{(k)})) L_d^{-H}) \\ &\leq \|(L_d^H L_d)\|_2^2 \operatorname{tr}(L_d^{-1} (X^{(k)} - X(\mu^{(k)})) L_d^{-H} L_d^{-1} (X^{(k)} - X(\mu^{(k)})) L_d^{-H}) \\ &\quad \text{par un raisonnement semblable au précédent} \\ &= \|L_d^H L_d\|_2^2 \|L_d^{-1} (X^{(k)} - X(\mu^{(k)})) L_d^{-H}\|_F^2. \end{aligned} \tag{4.31}$$

D'autre part, nous avons successivement

$$\begin{aligned} \|L_d^H L_d\|_F &= \|D_{\mu^{(k)}}^{-\frac{1}{2}} L_{\mu^{(k)}}^H L_{\mu^{(k)}} D_{\mu^{(k)}}^{-\frac{1}{2}}\|_F \\ &\quad \text{par définition de } L_\mu = L_d D_{\mu^{(k)}}^- \\ &\leq \|D_{\mu^{(k)}}^{-\frac{1}{2}}\|_2 \|L_{\mu^{(k)}}^H L_{\mu^{(k)}} D_{\mu^{(k)}}^{-\frac{1}{2}}\|_F \\ &\leq \|D_{\mu^{(k)}}^{-\frac{1}{2}}\|_2^2 \|L_{\mu^{(k)}}^H L_{\mu^{(k)}}\|_F. \end{aligned}$$

Or, par le lemme 4.4, $D_{\mu^{(k)}}^- = \Theta(1)$, c'est-à-dire

$$\begin{aligned} \exists \Gamma > 0 \quad \text{tel que} \quad \frac{1}{\Gamma} I &\leq D_{\mu^{(k)}}^- \leq \Gamma I \\ \frac{1}{\Gamma} I &\leq d_{\mu^{(k)},i}^- \leq \Gamma I \quad i = 1 \dots \bar{n} \end{aligned}$$

où $d_{\mu^{(k)},i}^-$ est la $i^{\text{ème}}$ valeur propre de $D_{\mu^{(k)}}^{-1}$. Il suit que $\|L_d^H L_d\|_F = O(\|L_{\mu^{(k)}}^H L_{\mu^{(k)}}\|_F)$. Par ailleurs, nous écrivons successivement

$$\begin{aligned} \|L_{\mu^{(k)}}^H L_{\mu^{(k)}}\|_F^2 &= L_{\mu^{(k)}}^H L_{\mu^{(k)}} \bullet L_{\mu^{(k)}}^H L_{\mu^{(k)}} \\ &= (X(\mu^{(k)}) \bullet X(\mu^{(k)})) \frac{1}{\mu^{(k)}} \\ &= \frac{1}{\mu^{(k)}} \|X(\mu^{(k)})\|_F^2 \end{aligned}$$

puisque $\sqrt{\mu^{(k)}} I = L_{\mu^{(k)}}^{-1} X(\mu^{(k)}) L_{\mu^{(k)}}^{-H}$ et donc $L_{\mu^{(k)}} L_{\mu^{(k)}}^H = \frac{X(\mu^{(k)})}{\sqrt{\mu^{(k)}}}$.

Par conséquent, nous obtenons

$$\begin{aligned} \|L_d^H L_d\|_F &= O\left(\frac{1}{\sqrt{\mu^{(k)}}} \|X(\mu^{(k)})\|_F\right) \\ &= O\left(\frac{1}{\sqrt{\mu^{(k)}}}\right) \end{aligned}$$

par le lemme 4.6. Nous obtenons dès lors, par (4.31)

$$\begin{aligned} \|X^{(k)} - X(\mu^{(k)})\|_F &\leq O\left(\frac{1}{\mu^{(k)}}\right) \|L_d^{-1}(X^{(k)} - X(\mu^{(k)})) L_d^{-H}\|_F^2 \\ &= O(\delta(V^{(k)})^2) \text{ par le lemme 4.3} \\ &= O(\mu^{(k)}) \end{aligned}$$

par 4.21. Par l'inégalité ci-dessus et par le théorème 4.2, nous avons

$$\|X^{(k)} - X^a\|_F \leq \|X^{(k)} - X(\mu^{(k)})\|_F + \|X(\mu^{(k)}) - X^a\|_F = O(\mu^{(k)}).$$

Nous concluons que $X^{(k)}$ converge vers le centre analytique X^a de l'ensemble des solutions optimales primales de la même manière que $\mu^{(k)}$.

De façon similaire, nous obtenons $\|Z^{(k)} - Z^a\|_F = O(\mu^{(k)})$. Nous pouvons également conclure que $Z^{(k)}$ converge vers le centre analytique Z^a de l'ensemble des solutions optimales duales de façon identique à $\mu^{(k)}$.

cqfd

Chapitre 5

Recherche d'informations sur le caractère admissible du problème ainsi que d'un itéré de départ intérieur dans le cas fortement admissible

La méthode primale-duale de points intérieurs et de suivi de chemin élaborée dans les chapitres précédents pour la résolution d'un problème de programmation semi-définie présuppose l'admissibilité forte primale et duale ainsi que la connaissance d'un point de départ strictement admissible. Dans la pratique cependant, à la vue d'un problème primal-dual, nous ne sommes en général pas capable de juger de son admissibilité. De plus, il n'est pas toujours évident de trouver un point de départ intérieur. Dans ce chapitre, nous résolvons, dans une certaine mesure, ces deux questions.

Nous débutons par l'étude de la self-dualité en programmation conique convexe. Celle-ci ainsi que la théorie de la dualité évoquée dans le chapitre 2 nous seront d'un grand secours, que ce soit pour déterminer un point de départ strictement admissible ou pour obtenir des informations sur le caractère admissible des problèmes primal et dual initiaux. Ensuite, en ce qui concerne la détermination d'un point strictement admissible, nous développons deux méthodes, généralisées de la programmation linéaire. La première est basée sur la self-dualité en programmation conique convexe et est applicable à tout problème de programmation conique convexe dont le cône convexe est fermé et solide. Elle reste donc valable pour la programmation semi-définie. La seconde, généralisée à la programmation semi-définie, est basée sur la technique du grand M.

Nous terminons par la recherche du caractère admissible des problèmes primal et dual grâce à l'existence de suites faiblement centrées, existence basée également sur la self-dualité.

Nous nous basons, tout au long de ce chapitre, sur l'écriture vectorielle, (1.6), d'un problème de programmation conique convexe.

5.1 La self-dualité

Nous définissons dans cette section la propriété de self-dualité en programmation conique convexe. Celle-ci permettra, dans les sections suivantes, de trouver un point initial $(X^{(0)}, Z^{(0)})$ intérieur, dans le cas fortement admissible, et de déterminer le caractère admissible ou non des problèmes primal et dual de départ. Nous élaborons ensuite quelques caractéristiques de la programmation conique convexe self-duale. Celles-ci faciliteront les démonstrations de ce chapitre.

Un problème de programmation conique convexe $CP(b, c, \mathcal{A}, \mathcal{K})$ est Π self-dual si et seulement si Π est une matrice de permutation symétrique telle que $c = \Pi b$, $\mathcal{A}^\perp = \Pi \mathcal{A}$, $\mathcal{K}^* = \Pi \mathcal{K}$.

Nous retrouvons la propriété d'égalité des problèmes primal et dual définissant la self-dualité en programmation linéaire. En effet, le dual est identique au primal après une simple réorganisation des variables. Dès lors, seule la diagonale du tableau (2.2) reste valable.

Notons également que la condition $\mathcal{K}^* = \Pi \mathcal{K}$ exige que \mathcal{K} soit un cône fermé, \mathcal{K}^* étant toujours fermé. Dès lors, les problèmes de programmation coniques convexes self-duaux sont fermés.

Etablissons quelques caractéristiques d'un problème de programmation conique convexe self-dual. Le plus important est celui de la dualité faible.

Théorème 5.1 *Si $CP(b, c, \mathcal{A}, \mathcal{K})$ est un problème de programmation semi-définie Π self-dual, alors $c^T x = \frac{1}{2} x^T \Pi x \geq 0$, $\forall x \in (b + \mathcal{A}) \cap \mathcal{K}$.*

preuve : Par hypothèse, nous avons que $x - b \in \mathcal{A}$, $\Pi x - c \in \mathcal{A}^\perp$ et $b \perp c$. Par conséquent, nous avons $0 = (x - b)^T (\Pi x - c) = x^T \Pi x - b^T \Pi x - c^T x$. Puisque $CP(b, c, \mathcal{A}, \mathcal{K})$ est self-dual, $c^T = b^T \Pi$. Dès lors, par la relation précédente, nous obtenons $2c^T x = x^T \Pi x$. Dès lors, comme $x \in \mathcal{K}$ et $\Pi x \in \mathcal{K}^*$, nous avons $c^T x = \frac{x^T \Pi x}{2} \geq 0$.

cqfd

D'une part, le lien avec le résultat de la dualité faible rencontré au chapitre 2 est aisé à effectuer. En effet, celui-ci affirme que $p^* + d^* \geq 0$ où p^* et d^* représentent les valeurs optimales primale et duale respectivement. Or, nous nous situons dans le cas où le problème de programmation conique convexe est self-dual. Dès lors, ce résultat devient $d^* = p^* \geq 0$. En d'autres mots, $c^T x = b^T \Pi x \geq c^T x^* = b^T \Pi x^* \geq 0$. Ceci implique en particulier qu'un problème de programmation conique convexe est borné.

D'autre part, une solution x^* est dite self-complémentaire si et seulement si x^* est un point admissible du problème $CP(b, c, \mathcal{A}, \mathcal{K})$ Π self-dual et $x^{*T} \Pi x^* = 0$. Le théorème 5.1 conclut donc que si x^* est une solution self-complémentaire alors x^* est une solution optimale. Ceci n'est pas négligeable et sera utilisée ultérieurement.

Les deuxième et troisième résultats élémentaires pour la programmation conique convexe self-duale sont les suivants :

Lemme 5.1 Soit \mathcal{A} un sous-espace linéaire de C^n tel que $\mathcal{A}^\perp = \Pi\mathcal{A}$ pour une certaine matrice de permutation symétrique Π . Alors $\Pi\mathcal{P}_{\mathcal{A}} = \mathcal{P}_{\mathcal{A}^\perp}\Pi$ et $\Pi\mathcal{P}_{\mathcal{A}^\perp} = \mathcal{P}_{\mathcal{A}}\Pi$ où $\mathcal{P}_{\mathcal{A}}$ et $\mathcal{P}_{\mathcal{A}^\perp}$ sont les matrices de projection sur \mathcal{A} et \mathcal{A}^\perp respectivement.

preuve : Soit $x \in C^n$. Sa décomposition orthogonale $x = y + z$ où $y \in \mathcal{A}$ et $z \in \mathcal{A}^\perp$ est équivalente à $\Pi x = \Pi y + \Pi z$ où $\Pi y \in \mathcal{A}^\perp$ et $\Pi z \in \mathcal{A}$. En effet, $\Pi\mathcal{A} = \mathcal{A}^\perp$ par hypothèse et $\Pi\mathcal{A}^\perp = \Pi(\Pi\mathcal{A}) = \mathcal{A}$ puisque $\Pi^T = \Pi^{-1}$, Π étant une matrice de permutation, et $\Pi = \Pi^T$, Π étant une matrice symétrique. Π est en fait orthogonale.

Par conséquent, nous obtenons, tout d'abord, $\Pi(\mathcal{P}_{\mathcal{A}}x) = \Pi y = \mathcal{P}_{\mathcal{A}^\perp}(\Pi x)$. Nous concluons à $\Pi\mathcal{P}_{\mathcal{A}} = \mathcal{P}_{\mathcal{A}^\perp}\Pi$, x étant choisi arbitrairement dans R^n .

Ensuite, nous avons $\Pi(\mathcal{P}_{\mathcal{A}^\perp}x) = \Pi z = \mathcal{P}_{\mathcal{A}}(\Pi x)$. Nous concluons à $\Pi\mathcal{P}_{\mathcal{A}^\perp} = \mathcal{P}_{\mathcal{A}}\Pi$, x étant choisi arbitrairement dans R^n .

cqfd

Lemme 5.2 Soit $CP(b, c, \mathcal{A}, \mathcal{K})$ un problème Π self-dual où \mathcal{K} est un cône solide convexe. Alors $y^T \mathcal{P}_{\mathcal{A}} \Pi y = y^T \mathcal{P}_{\mathcal{A}^\perp} \Pi y = \frac{1}{2} y^T \Pi y > 0 \quad \forall y \in \text{int}\mathcal{K}$.

preuve : \mathcal{K} étant solide, choisissons y à l'intérieur de \mathcal{K} .

Nous prouvons d'abord que $y^T \Pi y > 0$. Dans ce but, désirant appliquer le théorème 2.2, nous vérifions les hypothèses correspondantes.

Nous avons d'une part, $y \in \text{int}\mathcal{K}$.

D'autre part, vérifions que $\Pi y \in \mathcal{K}^*$. Nous savons que $y \in \text{int}\mathcal{K}$. Appliquant la matrice Π , nous obtenons $\Pi y \in \Pi \text{int}\mathcal{K}$. Or, nous avons $\Pi \text{int}\mathcal{K} = \text{int}\Pi\mathcal{K}$.

En effet, d'une part, si $x \in \text{int}\mathcal{K}$, $\exists r > 0$, tel que $B(x, r) \subseteq \mathcal{K}$. Donc, nous avons $\Pi B(x, r) \subseteq \Pi\mathcal{K}$. Montrons que $\Pi x \in \text{int}\Pi\mathcal{K}$, c'est-à-dire qu'il existe $\bar{r} > 0$ tel que $B(\Pi x, \bar{r}) \subseteq \Pi B(x, r) \subseteq \Pi\mathcal{K}$. Soit $y \in B(\Pi x, \bar{r})$. Nous avons successivement

$$\begin{aligned} \|y - \Pi x\| &< \bar{r} \\ \|\Pi(\Pi^{-1}y - x)\| &< \bar{r} \\ \|\Pi^{-1}y - x\| &< \bar{r} \end{aligned}$$

puisque Π est orthogonale. Il suit que $\Pi^{-1}y \in B(x, \bar{r})$. En conclusion, en choisissant $r = \bar{r}$, nous avons $y \in \Pi B(x, r)$.

D'autre part, si $y \in \text{int}\Pi\mathcal{K}$, il existe $r > 0$ tel que $B(y, r) \subseteq \Pi\mathcal{K}$. Donc, nous avons $\Pi^{-1}B(y, r) \subseteq \mathcal{K}$. Montrons que $y \in \Pi \text{int}\mathcal{K}$, c'est-à-dire que $\Pi^{-1}y \in \text{int}\mathcal{K}$. Prouvons qu'il existe $\bar{r} > 0$ tel que $B(\Pi^{-1}y, \bar{r}) \subseteq \Pi^{-1}B(y, r) \subseteq \mathcal{K}$. Soit $x \in B(\Pi^{-1}y, \bar{r})$. Nous écrivons successivement

$$\begin{aligned} \|x - \Pi^{-1}y\| &< \bar{r} \\ \|\Pi^{-1}(\Pi x - y)\| &< \bar{r} \\ \|\Pi x - y\| &< \bar{r}. \end{aligned}$$

Il suit que $\Pi x \in B(y, \bar{r})$. Choisissons $r = \bar{r}$, nous obtenons $x \in \Pi^{-1}B(y, r)$.

Par conséquent, $\Pi y \in \text{int}\Pi\mathcal{K} = \text{int}\mathcal{K}^*$ par la définition d'un problème self-dual.

La troisième hypothèse à vérifier est $\Pi y \notin \mathcal{K}^\perp$. Celle-ci est satisfaite puisque, si nous supposons le contraire, comme $\text{int}\mathcal{K} \neq \emptyset$, nous avons $\Pi y = 0$. Or, Πy doit être différent de 0. En effet, $0 \notin \text{int}\mathcal{K}^*$ puisque, \mathcal{K} étant solide, \mathcal{K}^* est pointé par le corollaire 2.1.

Dès lors, appliquant le théorème 2.2, nous obtenons $0 < y^T \Pi y$.

Terminons cette preuve par la démonstration des égalités restantes. Nous avons en décomposant y ,

$$y^T \Pi y = y^T \Pi (\mathcal{P}_{\mathcal{A}} y + \mathcal{P}_{\mathcal{A}^\perp} y) = y^T \Pi \mathcal{P}_{\mathcal{A}} y + y^T \Pi \mathcal{P}_{\mathcal{A}^\perp} y.$$

D'une part, nous avons $y^T \Pi \mathcal{P}_{\mathcal{A}^\perp} y = y^T \mathcal{P}_{\mathcal{A}^\perp}^T \Pi^T y = y^T \mathcal{P}_{\mathcal{A}^\perp} \Pi y$, car Π et $\mathcal{P}_{\mathcal{A}^\perp}$ sont des matrices symétriques. Nous obtenons ainsi $0 < y^T \Pi y = 2y^T \mathcal{P}_{\mathcal{A}^\perp} \Pi y$ par le lemme 5.1.

D'autre part, nous avons $y^T \Pi \mathcal{P}_{\mathcal{A}} y = y^T \mathcal{P}_{\mathcal{A}}^T \Pi^T y = y^T \mathcal{P}_{\mathcal{A}} \Pi y$. Nous obtenons $0 < y^T \Pi y = 2y^T \mathcal{P}_{\mathcal{A}} \Pi y$, par un raisonnement similaire.

cqfd

Nous pouvons dès lors nous attacher à résoudre la question de la recherche d'un point de départ strictement admissible et celle de la recherche du caractère admissible des problèmes primal et dual.

5.2 Recherche d'un point intérieur initial $(x^{(0)}, z^{(0)})$

Sous l'hypothèse de forte admissibilité, les algorithmes détaillés dans les chapitres 3 et 4 présupposent tous l'existence d'un point intérieur de départ $(x^{(0)}, z^{(0)})$. Dans la pratique cependant, celui-ci n'est pas toujours facile à calculer. Dans cette section, nous tâchons de trouver un tel point de départ. Nous le faisons en généralisant deux méthodes utilisées en programmation linéaire.

Nous généralisons d'abord, à la programmation conique convexe, la méthode de la programmation linéaire basée sur la self-dualité. Pour se faciliter la tâche, nous supposons en premier lieu le problème de programmation conique convexe initial self-dual. Ensuite, nous étudions le cas où le problème initial est non nécessairement self-dual. Cette étude exige l'hypothèse d'un cône solide et fermé. Puisque le cône semi-défini positif vérifie ces deux caractéristiques, cette méthode reste valable en programmation semi-définie.

Nous généralisons ensuite la méthode du grand M à la programmation semi-définie. Celle-ci se base sur l'introduction dans le problème initial d'une très grande constante positive M.

5.2.1 Méthode basée sur la self-dualité : cas d'un problème de programmation conique convexe self-dual

Considérons un problème de programmation conique convexe fortement admissible et Π_{SD} self-dual, $CP(b_{SD}, c_{SD}, \mathcal{A}_{SD}, \mathcal{K}_{SD})$:

$$(SD) \quad \inf \{ c_{SD}^T x_{SD} \mid x_{SD} \in (b_{SD} + \mathcal{A}_{SD}) \cap \mathcal{K}_{SD} \} \quad (5.1)$$

où \mathcal{K}_{SD} est solide. Nous faisons cette hypothèse afin de pouvoir utiliser le lemme 5.2. Nous notons la valeur optimale de (SD), p_{SD}^* .

Nous avons par hypothèse la forte admissibilité de (SD). Néanmoins, nous ne connaissons pas à priori de point strictement admissible pouvant servir d'itéré de départ pour une méthode de points intérieurs.

Pour contourner cette difficulté, nous allons construire, sur base du problème self-dual (SD), deux problèmes de programmation conique convexe notés (H) et (E). Nous prouverons, grâce aux propriétés du problème (H) qu'il est aisé de trouver un point strictement admissible pour (E). Nous pourrons dès lors résoudre ce problème par une méthode de points intérieurs nécessitant la connaissance d'un itéré de départ strictement admissible. Nous démontrerons ensuite qu'à toute solution self-complémentaire de (E), et donc optimale, correspond une solution self-complémentaire de (SD), et donc optimale. Il s'avèrera que celle-ci est aisément calculable à partir de la solution optimale de (E) trouvée.

La construction des problèmes (H) et (E) est la généralisation de la programmation linéaire à la programmation conique convexe des problèmes (HSD) et (mLCP) de Ye, Todd et Mizuno.

construction du problème de programmation conique convexe self-dual homogène (H) associé à (SD)

Nous définissons ici le problème (H). Ensuite, nous dégageons certaines de ses propriétés et démontrons un résultat indispensable pour la suite.

Nous introduisons $M(b_{SD}, c_{SD}) := \begin{pmatrix} I & b_{SD} & 0 \\ 0 & 1 & 0 \\ -c_{SD}^T & 0 & 1 \end{pmatrix}$ une matrice inversible. Nous définissons le nouveau problème self-dual homogène (H) comme

$$(H) : \quad CP(0, 0, \mathcal{A}_H, \mathcal{K}_H) \quad (5.2)$$

où $\mathcal{A}_H := M(b_{SD}, c_{SD})(\mathcal{A}_{SD} \times R \times \{0\})$ et $\mathcal{K}_H = \mathcal{K}_{SD} \times R^+ \times R^+$.

Nous remarquons que (H) est Π_H self-dual avec

$$\Pi_H = \begin{pmatrix} \Pi_{SD} & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \quad (5.3)$$

En effet, Π_H est une matrice de permutation symétrique car Π_{SD} en est une. De plus, nous avons $c_H = 0 = \Pi_H b_H$, $\Pi_H \mathcal{K}_H = \mathcal{K}_H^*$. Finalement, nous obtenons $\Pi_H \mathcal{A}_H = \mathcal{A}_H^\perp$. En effet, nous avons

$$\Pi_H \mathcal{A}_H = \begin{pmatrix} \Pi_{SD} & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathcal{A}_{SD} + b_{SD} \cdot R \\ R \\ -c_{SD}^T \mathcal{A}_{SD} + \{0\} \end{pmatrix}$$

$$= (\mathcal{A}_{SD}^\perp + c_{SD} \cdot R) \times (-c_{SD}^T \mathcal{A}_{SD} + \{0\}) \times R.$$

Or, nous avons $\mathcal{A}_H^\perp = M(b_{SD}, c_{SD})^{-T}(\mathcal{A}_{SD}^\perp \times \{0\} \times R)$ par le lemme 2.2. Donc, nous obtenons

$$\begin{aligned} \mathcal{A}_H^\perp &= \begin{pmatrix} I & 0 & c_{SD} \\ -b_{SD}^T & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathcal{A}_{SD}^\perp \\ \{0\} \\ R \end{pmatrix} \\ &= (\mathcal{A}_{SD}^\perp + c_{SD} \cdot R) \times (-b_{SD}^T \mathcal{A}_{SD}^\perp + \{0\}) \times R. \end{aligned}$$

Tenant compte du fait que $b_{SD}^T \mathcal{A}_{SD}^\perp = c_{SD}^T \mathcal{A}_{SD}$, nous obtenons $\Pi_H \mathcal{A}_H = \mathcal{A}_H^\perp$.

Nous remarquons également que (H) est un problème homogène puisque sa fonction objectif est nulle. Par conséquent, les solutions optimales du problème (H) sont les solutions admissibles de ce problème. Elles sont donc les points (x_{SD}, x_0, z_0) qui appartiennent à $\mathcal{A}_H \cap \mathcal{K}_H$, ce qui est équivalent à dire qu'elles sont les solutions du système suivant :

$$\begin{cases} x_{SD} \in (x_0 b_{SD} + \mathcal{A}_{SD}) \cap \mathcal{K}_{SD} \\ x_0 \geq 0 \\ z_0 = -c_{SD}^T x_{SD} \geq 0 \end{cases} \quad (5.4)$$

Ces observations nous permettent de démontrer le théorème suivant. Celui-ci fournit la condition nécessaire que (H) doit satisfaire pour obtenir une solution optimale du problème (SD) de programmation conique convexe self-dual de départ.

Théorème 5.2 Si $(x_{SD}, x_0, z_0) \in \mathcal{A}_H \cap \mathcal{K}_H$, alors $x_{SD}^T \Pi_{SD} x_{SD} = 0$ et $x_0 z_0 = 0$.

De plus, si $(x_{SD}, x_0, z_0) \in \mathcal{A}_H \cap \mathcal{K}_H \setminus \{0\}$ et si (SD) est fortement admissible, alors $x_0 > 0$ et $\frac{x_{SD}}{x_0}$ est une solution self-complémentaire de (SD).

preuve : Supposons tout d'abord que $(x_{SD}, x_0, z_0) \in \mathcal{A}_H \cap \mathcal{K}_H$ et prouvons que $x_{SD}^T \Pi_{SD} x_{SD} = 0$ ainsi que $x_0 z_0 = 0$.

Puisque (x_{SD}, x_0, z_0) est une solution admissible de (H), celle-ci résout le système (5.4). De la dernière équation, nous avons $z_0 = -c_{SD}^T x_{SD} \geq 0$. Par le théorème 5.1, nous avons $c_{SD}^T x_{SD} = \frac{1}{2} x_{SD}^T \Pi_{SD} x_{SD} \geq 0$. Nous obtenons ainsi $z_0 = -\frac{1}{2} x_{SD}^T \Pi_{SD} x_{SD} = 0$ et concluons que $x_0 z_0 = 0$ et $x_{SD}^T \Pi_{SD} x_{SD} = 0$.

Ensuite, supposons que $(x_{SD}, x_0, z_0) \in (\mathcal{A}_H \cap \mathcal{K}_H) \setminus \{0\}$ et que (SD) est fortement admissible. Prouvons alors que $x_0 > 0$ et que $\frac{x_{SD}}{x_0}$ est une solution self-complémentaire de (SD).

D'une part, (5.4) impose à x_0 d'être positif. Supposons par la contraposée x_0 nul. Nous avons que $(x_{SD}, 0, z_0) \in \mathcal{A}_H \cap \mathcal{K}_H$ où on suppose $z_0 > 0$ pour satisfaire l'hypothèse de non-nullité de (x_{SD}, x_0, z_0) . Ceci est équivalent à dire que x_{SD} est une direction améliorante pour (SD). Cette équivalence est obtenue en utilisant (5.4). Par le tableau 2.1, ceci revient à dire que (SD) est fortement inadmissible, ce qui est impossible. Dès lors, nous obtenons que x_0 doit être strictement positif.

D'autre part, nous avons l'équivalence suivante

$$y \text{ est une solution self-complémentaire de (SD)} \iff (y, 1, 0) \in \mathcal{A}_H \cap \mathcal{K}_H. \quad (5.5)$$

En effet, y est une solution self-complémentaire de (SD) si et seulement si

$$\begin{cases} y \in (b_{SD} + \mathcal{A}_{SD}) \cap \mathcal{K}_{SD} \\ y^T \Pi_{SD} y = 0. \end{cases}$$

Ce système est équivalent, après quelques transformations, à celui-ci

$$\begin{cases} y \in (b_{SD} + \mathcal{A}_{SD}) \cap \mathcal{K}_{SD} \\ x_0 = 1 \geq 0 \\ z_0 = -c_{SD}^T y = -\frac{1}{2} y^T \Pi_{SD} y = 0. \end{cases}$$

Par (5.4), nous avons que $(y, 1, 0)$ est un élément de $\mathcal{A}_H \cap \mathcal{K}_H$.

Il nous reste à montrer que $(\frac{x_{SD}}{x_0}, 1, 0) \in \mathcal{A}_H \cap \mathcal{K}_H$ pour obtenir la thèse. Dès lors, vérifions que $(\frac{x_{SD}}{x_0}, 1, 0)$ est une solution de (5.4), c'est-à-dire que

$$\begin{cases} \frac{x_{SD}}{x_0} \in (b_{SD} + \mathcal{A}_{SD}) \cap \mathcal{K}_{SD} \\ 1 \geq 0 \\ 0 = -c_{SD}^T \frac{x_{SD}}{x_0}. \end{cases}$$

Or, nous savons que $\frac{x_{SD}}{x_0} \in (b_{SD} + \mathcal{A}_{SD}) \cap \mathcal{K}_{SD}$ puisque $(x_{SD}, x_0, z_0) \in \mathcal{A}_H \cap \mathcal{K}_H$ et $x_0 > 0$. De plus, $-c_{SD}^T x_{SD} = 0$, par un raisonnement identique à celui du début de la démonstration. Par conséquent, $\frac{x_{SD}}{x_0}$ est une solution self-complémentaire de (SD).

cqfd

Ce théorème nous permet de conclure que, sous l'hypothèse de forte admissibilité de (SD), si nous trouvons une solution admissible non-nulle pour (H), soit (x_{SD}, x_0, z_0) , alors $\frac{x_{SD}}{x_0}$ est une solution self-complémentaire de (SD) et donc optimale pour ce problème. Ce résultat sera utilisé par la suite, puisqu'à partir d'une solution optimale de (H), nous trouvons une solution optimale de (SD).

Construction du problème de programmation conique convexe self-dual étendu (E) associé à (SD)

Nous avons associé au problème self-dual fortement admissible (SD) un problème self-dual homogène (H). Nous allons à présent construire un autre problème (E), qui est fortement admissible et dont la connaissance d'un itéré de départ strictement admissible ne pose pas de problèmes.

Nous définissons d'abord ce nouveau problème (E), ensuite nous prouvons qu'il est self-dual. Pour terminer, nous formulons ce problème d'une autre manière. Celle-ci nous permettra aisément de démontrer la forte admissibilité de (E).

Définissons dans un premier temps les paramètres nécessaires à la construction du problème de programmation conique convexe self-dual étendu,

$$(E) : \quad CP(b_E, c_E, \mathcal{A}_E, \mathcal{K}_E). \quad (5.6)$$

Choisissons arbitrairement

$$i \in \text{int} \mathcal{K}_H \quad (5.7)$$

et posons

$$\rho := \frac{i^T \mathcal{P}_{\mathcal{A}_H} \Pi_H i}{\|\mathcal{P}_{\mathcal{A}_H^\perp} i\|_2^2}. \quad (5.8)$$

Ainsi, ρ est strictement positif, puisque, vu que \mathcal{K}_H est solide, nous appliquons le lemme 5.2 et obtenons

$$i^T \mathcal{P}_{\mathcal{A}_H} \Pi_H i = i^T \mathcal{P}_{\mathcal{A}_H^\perp} \Pi_H i = \frac{1}{2} i^T \Pi_H i > 0. \quad (5.9)$$

Définissons alors b_E, c_E, \mathcal{A}_E et \mathcal{K}_E .

Nous posons

$$b_E := \rho \mathcal{P}_{\mathcal{A}_H} \Pi_H i. \quad (5.10)$$

Puisque nous imposons à (E) d'être Π_H self-dual, nous avons

$$c_E = \Pi_H b_E = \Pi_H \rho \mathcal{P}_{\mathcal{A}_H} \Pi_H i = \rho \mathcal{P}_{\mathcal{A}_H^\perp} i \quad (5.11)$$

par le lemme 5.1. Nous voyons ici que (E) n'est pas un problème de programmation conique convexe homogène. En effet, nous avons successivement

$$\begin{aligned} \frac{\|b_E\|_2^2}{\rho} &= \frac{\|c_E\|_2^2}{\rho} \quad c_E \text{ étant une permutation de } b_E \\ &= \rho \|\mathcal{P}_{\mathcal{A}_H^\perp} i\|_2^2 \\ &= i^T \mathcal{P}_{\mathcal{A}_H} \Pi_H i \text{ par définition de } \rho \\ \frac{\|b_E\|_2^2}{\rho} &= \frac{i^T \Pi_H i}{2} > 0 \text{ par (5.9).} \end{aligned} \quad (5.12)$$

Posons également,

$$\mathcal{A}_E := (\mathcal{A}_H \cap \ker b_E^T) \oplus \text{im } c_E \quad \text{et} \quad \mathcal{K}_E := \mathcal{K}_H. \quad (5.13)$$

Dans un second temps, prouvons que le problème $CP(b_E, c_E, \mathcal{A}_E, \mathcal{K}_E)$ est Π_H self-dual. Nous savons déjà que Π_H est une matrice de permutation symétrique, que $c_E = \Pi_H b_E$ par (5.11) et que $\mathcal{K}_E^* = \mathcal{K}_H^* = \Pi_H \mathcal{K}_H = \Pi_H \mathcal{K}_E$ par self-dualité du problème (H). Il reste donc à montrer que $\mathcal{A}_E^\perp = \Pi_H \mathcal{A}_E$. Nous avons successivement

$$\begin{aligned} \mathcal{A}_E^\perp &= [(\mathcal{A}_H \cap \ker b_E^T) \oplus \text{im } c_E]^\perp \\ &= (\mathcal{A}_H \cap \ker b_E^T)^\perp \cap (\text{im } c_E)^\perp \quad \text{par le lemme 2.1.} \end{aligned}$$

Or, nous savons que $(\text{im } c_E)^\perp = \ker c_E^T$ et $(\mathcal{A}_H \cap \ker b_E^T)^\perp = ((\mathcal{A}_H)^\perp \oplus \text{im } b_E)^\perp$ par le lemme 2.1. Par conséquent, nous écrivons

$$\begin{aligned} \mathcal{A}_E^\perp &= ((\mathcal{A}_H)^\perp \oplus \text{im } b_E) \cap \ker c_E^T \\ &= (\mathcal{A}_H^\perp \cap \ker c_E^T) \oplus \text{im } b_E \\ &= (\Pi_H \mathcal{A}_H \cap \ker b_E^T \Pi_H) \oplus \text{im } \Pi_H c_E \\ &= \Pi_H [(\mathcal{A}_H \cap \ker b_E^T) \oplus \text{im } c_E] \\ &= \Pi_H \mathcal{A}_E. \end{aligned}$$

Dans un troisième temps, formulons (E) autrement grâce à l'introduction d'une variable auxiliaire

$$y_0 := \frac{\rho}{\|c_E\|_2^2} c_E^T x_E.$$

Cette nouvelle expression de (E) nous permettra de prouver plus facilement que tout point i choisi à l'intérieur de \mathcal{K}_H est un point strictement admissible de (E). Nous aurons ainsi un point de départ intérieur pour (E).

Tout d'abord, intéressons-nous à la fonction objectif de (E). Nous avons successivement

$$\begin{aligned} c_E^T x_E &= \frac{\rho}{\|c_E\|_2^2} c_E^T x_E \frac{\|c_E\|_2^2}{\rho} \\ &= y_0 \frac{i^T \Pi_H i}{2} \end{aligned} \quad (5.14)$$

par (5.12).

Examinons ensuite la contrainte. Nous obtenons successivement

$$\begin{aligned} x_E \in (b_E + \mathcal{A}_E) &\iff (x_E - b_E) \in (\mathcal{A}_H \cap \ker b_E^T) \oplus \text{im } c_E \\ &\iff x_E - b_E - y c_E \in \mathcal{A}_H \cap \ker b_E^T \quad \text{où } y \in R \\ &\iff \begin{cases} x_E - b_E - y c_E \in \mathcal{A}_H \\ b_E^T x_E = \|b_E\|_2^2 \end{cases} \end{aligned}$$

puisque b_E est orthogonal à c_E . Or, b_E appartient à \mathcal{A}_H par définition. Par conséquent, $x_E - b_E - y c_E \in \mathcal{A}_H$ est équivalent à $x_E - y c_E \in \mathcal{A}_H$. De plus, nous avons $c_E \in \mathcal{A}_H^\perp$ par définition. Nous avons donc $c_E^T x_E - \|c_E\|_2^2 y = 0$, ce qui veut dire que y vaut $\frac{c_E^T x_E}{\|c_E\|_2^2}$. Dès lors, la contrainte, $x_E - b_E - y c_E \in \mathcal{A}_H$, devient successivement

$$\begin{aligned} &x_E - \frac{c_E^T x_E}{\|c_E\|_2^2} c_E \in \mathcal{A}_H \\ \iff &x_E - \frac{y_0}{\rho} c_E \in \mathcal{A}_H \\ \iff &x_E - y_0 \mathcal{P}_{\mathcal{A}_H^\perp} i \in \mathcal{A}_H \\ \iff &x_E - y_0 i \in \mathcal{A}_H. \end{aligned}$$

Par conséquent, le problème de programmation conique convexe self-dual étendu associé à (SD) peut s'écrire comme

$$\begin{aligned} &\min y_0 \frac{i^T \Pi_H i}{2} \\ \text{s.c. } &\begin{cases} x_E - y_0 i \in \mathcal{A}_H \\ b_E^T x_E = \|b_E\|_2^2 \\ x_E \in \mathcal{K}_H \\ y_0 \in R \end{cases} \end{aligned} \quad (5.15)$$

Grâce à cette nouvelle notation, nous allons prouver plus aisément que

$$\forall i \in \text{int} \mathcal{K}_H, \quad i \in (b_E + \mathcal{A}_E). \quad (5.16)$$

Ainsi, tout $i \in \text{int}\mathcal{K}_H$ est un point strictement admissible de (E). En effet, démontrons cette relation. Nous devons montrer que $\forall i \in \text{int}\mathcal{K}_H$,

$$\begin{cases} i - y_0 i \in \mathcal{A}_H \\ b_E^T i = \|b_E\|_2^2 \\ y_0 \in R. \end{cases} \quad (5.17)$$

Examinons la contrainte de normalisation, $b_E^T x_E = \|b_E\|_2^2$, de manière générale. Nous savons que $b_E^T x_E = \|b_E\|_2^2 \iff \frac{b_E^T x_E}{\rho} = \frac{\|b_E\|_2^2}{\rho}$. Or, nous savons que d'une part, par (5.12), $\frac{\|b_E\|_2^2}{\rho} = \frac{i^T \Pi_H i}{2}$ et que d'autre part

$$\begin{aligned} \frac{b_E^T x_E}{\rho} &= (\mathcal{P}_{\mathcal{A}_H} \Pi_H i)^T x_E \text{ par (5.10)} \\ &= i^T \Pi_H \mathcal{P}_{\mathcal{A}_H} x_E \text{ par symétrie de } \Pi_H \text{ et } \mathcal{P}_{\mathcal{A}_H} \\ &= i^T \Pi_H (x_E - \mathcal{P}_{\mathcal{A}_H^\perp} x_E) \\ &= i^T \Pi_H x_E - i^T \Pi_H \mathcal{P}_{\mathcal{A}_H^\perp} x_E \\ &= i^T \Pi_H x_E - i^T \Pi_H (y_0 \mathcal{P}_{\mathcal{A}_H^\perp} i) \text{ puisque } x_E \in y_0 i + \mathcal{A}_H \\ &= i^T \Pi_H x_E - \frac{y_0}{2} i^T \Pi_H i \text{ par le lemme 5.2.} \end{aligned} \quad (5.18)$$

Donc, la contrainte de normalisation, $b_E^T x_E = \|b_E\|_2^2$, est satisfaite pour x_E si et seulement si

$$i^T \Pi_H x_E = \frac{1 + y_0}{2} i^T \Pi_H i. \quad (5.19)$$

Nous voyons immédiatement que pour $x_E = i$ la contrainte de normalisation est satisfaite pour $y_0 = 1$. Les deux autres équations sont également satisfaites avec $y_0 = 1$ et $x_E = i$.

Nous avons dès lors obtenu un résultat important : i appartient à l'intérieur de l'ensemble admissible de (E). i peut donc servir de point de départ pour les méthodes de points intérieurs pour la résolution du problème de programmation conique convexe self-dual étendu. Ce qui est quelque peu surprenant est que la seule condition imposée à i est d'appartenir à l'intérieur du cône \mathcal{K}_H . Heureusement, rechercher un tel point n'est pas trop compliqué.

Méthode de résolution d'un problème de programmation conique convexe self-dual (SD) fortement admissible ne nécessitant pas la connaissance d'un itéré de départ strictement admissible

Nous nous sommes rendus compte qu'un problème de programmation conique convexe self-dual fortement admissible noté (SD) peut être transformé en un problème de programmation conique convexe self-dual étendu (E) pour lequel il est aisé de trouver un itéré initial $(x_E^{(0)}, z_E^{(0)})$. Nous prouvons maintenant qu'à toute solution self-complémentaire de (E), $x_E = (x, x_0, z_0)$, qui est une solution optimale de (E), correspond une direction non

nulle du problème de programmation conique convexe homogène (H). Utilisant alors le théorème (5.2), sous l'hypothèse de forte admissibilité de (SD), nous trouvons qu'à toute solution self-complémentaire de (E), $x_E = (x, x_0, z_0)$, et donc optimale, correspond une solution self-complémentaire de (SD), et donc optimale.

Dès lors, nous concluons que pour obtenir une solution optimale pour (SD), avec (SD) fortement admissible, par une méthode de points intérieurs qui nécessite la connaissance d'un point initial intérieur, $(x^{(0)}, z^{(0)} = \Pi_{SD}x^{(0)})$, il suffit de résoudre le problème de programmation conique convexe self-dual étendu (E), qui correspond à (SD), par une méthode de points intérieurs. Celle-ci est applicable puisque nous pouvons choisir aisément un point initial admissible $(i, \Pi_E i)$ où $i \in \text{int}\mathcal{K}_H$.

Intéressons-nous donc à prouver qu'à toute solution self-complémentaire de (E) correspond une direction non-nulle de (H).

Théorème 5.3 *Sous les hypothèses faites sur (SD), (H) et (E), nous avons les deux assertions suivantes :*

- 1) *Toute solution self-complémentaire de $CP(b_E, c_E, \mathcal{A}_E, \mathcal{K}_E)$ est une direction non-nulle de $CP(0, 0, \mathcal{A}_H, \mathcal{K}_H)$.*
- 2) *Pour toute direction non-nulle x_E du problème de programmation conique convexe homogène $CP(0, 0, \mathcal{A}_H, \mathcal{K}_H)$, il existe α strictement positif tel que αx_E est une solution self-complémentaire de $CP(b_E, c_E, \mathcal{A}_E, \mathcal{K}_E)$.*

preuve : 1) Soit x_E une solution self-complémentaire de (E). Une telle solution existe par le théorème 2.6 et le fait que (E) possède un point intérieur i . Par définition et par le théorème 5.1, nous avons que

$$0 = x_E^T \Pi_H x_E = 2c_E^T x_E = 2y_0 \frac{\|c_E\|_2^2}{\rho}.$$

Ainsi, nous avons par (5.12) que $y_0=0$. Nous pouvons donc conclure par (5.15) que $x_E \in \mathcal{A}_H \cap \mathcal{K}_H$. Il reste à montrer que x_E est différent de 0. En effet, par (5.12) et (5.19), nous avons $i^T \Pi_H x_E = \frac{1}{2} i^T \Pi_H i > 0$, inégalité qui ne peut être satisfaite pour $x_E = 0$.

2) Soit $x_E \in (\mathcal{A}_H \cap \mathcal{K}_H) \setminus \{0\}$. Tout d'abord, prouvons que $i^T \Pi_H x_E > 0$ en utilisant le théorème 2.2. Cette inégalité sera nécessaire pour la suite de la démonstration.

Par la preuve du lemme 5.2, nous avons $\Pi_H i \in \text{int}\mathcal{K}_H^* = \text{rel}\mathcal{K}_H^*$. Donc, nous obtenons que $x_E \in [(\mathcal{K}_H^*)^* \setminus (\mathcal{K}_H^*)^\perp]$ entraîne $(\Pi_H i)^T x_E = i^T \Pi_H x_E > 0$. Or, nous savons que $x_E \in \mathcal{K}_H = \text{cl}\mathcal{K}_H = (\mathcal{K}_H^*)^*$ par le théorème 2.1, et que $x_E \notin (\mathcal{K}_H^*)^\perp$. En effet, si nous supposons par contraposition que $x_E \in (\mathcal{K}_H^*)^\perp$, nous obtenons $x_E^T k = 0$ pour tout $k \in \mathcal{K}_H^*$. Or, nous avons $\Pi_H i \in \text{int}\mathcal{K}_H^*$. Donc il existe un $r > 0$ tel que $B(\Pi_H i, r) \subseteq \mathcal{K}_H^*$. Par conséquent, pour tout $k \in B(\Pi_H i, r)$, $x_E^T k = 0$. Un tel x_E ne peut être que nul, ce qui est impossible par hypothèse. Ainsi, nous obtenons $i^T \Pi_H x_E > 0$.

En prouvant ensuite que $x = \frac{i^T \Pi_H i}{2i^T \Pi_H x_E} x_E \in (b_E + \mathcal{A}_H) \cap \mathcal{K}_H$, nous obtenons que $x = \alpha x_E$, où $\alpha = \frac{i^T \Pi_H i}{2i^T \Pi_H x_E} > 0$, est une solution self-complémentaire, puisque x est admissible et $\alpha^2 x_E^T \Pi_H x_E = \alpha^2 (x_{SD}^T \Pi_{SD} x_{SD}) + 2x_0 z_0 = 0$ par le théorème 5.2 et par la forte admissibilité

de (SD).

Démontrons donc que $x = \alpha x_E \in (b_E + \mathcal{A}_H) \cap \mathcal{K}_H$.

Puisque α est strictement positif et $x_E \in \mathcal{K}_H$, nous avons que $\alpha x_E \in \mathcal{K}_H$.

Il reste à vérifier la contrainte $\alpha x_E \in (b_E + \mathcal{A}_H)$. Par (5.15), nous devons vérifier $\alpha x_E - y_0 i \in \mathcal{A}_H$ et $\alpha b_E^T x_E = \|b_E\|_2^2$. D'une part, par hypothèse $x_E \in \mathcal{A}_H$ et donc $\alpha x_E \in \mathcal{A}_H$. De plus, puisque $(\alpha x_E)^T \Pi_H(\alpha x_E) = 0$, nous avons $y_0 = 0$ par un raisonnement analogue à celui développé dans 1). Dès lors, nous avons $\alpha x_E - y_0 i \in \mathcal{A}_H$. D'autre part, par (5.18) et puisque $y_0 = 0$, nous avons successivement

$$\begin{aligned} \frac{b_E^T x_E}{\rho} &= i^T \Pi_H x_E \\ \frac{i^T \Pi_H i}{2i^T \Pi_H x_E} b_E^T x_E &= \rho \frac{i^T \Pi_H i}{2} = \|b_E\|_2^2 \end{aligned}$$

par (5.12). Ainsi, $\frac{i^T \Pi_H i}{2i^T \Pi_H x_E} x_E$ est une solution self-complémentaire de (E).

cqfd

Grâce à ce théorème, nous sommes à même de trouver une solution optimale d'un problème de programmation conique convexe self-dual (SD), fortement admissible et dont on ne connaît pas de point de départ strictement admissible, en résolvant par une méthode de points intérieurs le problème de programmation conique convexe self-dual étendu (E) associé, pour lequel il est aisé de trouver un point initial strictement admissible.

Cependant, ce résultat repose sur le caractère self-dual de (SD). Nous voulons à présent, étendre celui-ci à tout problème de programmation conique convexe fermée.

5.2.2 Méthode basée sur la self-dualité : extension au cas de la programmation conique convexe fermée

Nous allons montrer à présent que la technique développée précédemment est applicable à la programmation conique convexe fermée non nécessairement self-duale.

En effet, supposons que nous devons résoudre $CP(b, c, \mathcal{A}, \mathcal{K})$, ainsi que son dual. Supposons ceux-ci fortement admissibles. Ces deux-ci ne sont pas nécessairement identiques. Pour pouvoir utiliser les résultats précédents, nous nous ramenons à un problème de programmation conique convexe self-dual, (SD). Nous posons

$$\begin{aligned} x_{SD} &:= (x^T, z^T)^T \\ b_{SD} &:= [b^T, c^T]^T \\ c_{SD} &:= [c^T, b^T]^T \\ \mathcal{A}_{SD} &:= \mathcal{A} \times \mathcal{A}^\perp \\ \mathcal{K}_{SD} &:= \mathcal{K} \times \mathcal{K}^* \text{ où } \mathcal{K} \text{ est un cône fermé et solide} \\ \Pi_{SD} &= \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \end{aligned} \tag{5.20}$$

Nous remarquons que le problème (SD) n'est rien d'autre que le problème de programmation conique convexe primal-dual :

$$(SD) : \inf c^T x + b^T z \quad \text{s.c. } x \in (b + \mathcal{A}) \cap \mathcal{K} \text{ et } z \in (c + \mathcal{A}^\perp) \cap \mathcal{K}^*.$$

Nous voyons également que ce problème est Π_{SD} self-dual. Nous pouvons dès lors résoudre le problème de programmation conique convexe fermée (SD) fortement admissible en résolvant le problème de programmation conique convexe self-dual étendu associé (E). Celui-ci nous permet donc de trouver une solution optimale primale-duale.

Notons que nous imposons à \mathcal{K} d'être fermé et solide pour que \mathcal{K}_{SD} satisfasse aux hypothèses du point précédent. Remarquons également que dans le cas particulier de la programmation semi-définie, $\mathcal{K} = \mathcal{H}^+$ est un cône fermé et solide. Par conséquent, tout ceci reste valable.

5.2.3 La méthode du grand M

Nous développons ici pour la programmation semi-définie une autre méthode qui permet de trouver un itéré de départ d'un problème primal-dual fortement admissible, $(x^{(0)}, z^{(0)})$, la méthode du grand M. Celle-ci est une simple généralisation de la méthode du grand M de la programmation linéaire. Elle ne tient donc nullement compte du caractère self-dual du problème. En effet, elle se base sur l'introduction d'une très grande constante positive notée M. Dans ce point, nous utilisons la notation (1.1) du problème de programmation semi-définie, qui est, rappelons-la,

$$\min c^T x \quad \text{s.c. } F(x) \succeq 0.$$

Le dual de ce problème est

$$\max -\text{Retr} F_0 Z \quad \text{s.c. } \text{Retr} F_i Z = c_i \text{ pour } i = 1 \dots m, \quad Z \succeq 0.$$

Nous distinguons trois cas. Pour le premier, nous connaissons $x^{(0)}$ strictement admissible mais pas de $Z^{(0)}$ strictement admissible. En ce qui concerne le second, nous connaissons $Z^{(0)}$ strictement admissible mais pas de $x^{(0)}$ strictement admissible. Pour le troisième, nous ne connaissons ni $x^{(0)}$ strictement admissible ni $Z^{(0)}$ strictement admissible.

Regardons comment résoudre le problème de la recherche d'un itéré de départ $(x^{(0)}, Z^{(0)})$ dans chacun de ces trois cas par la méthode du grand M.

Connaissance d'un $x^{(0)}$ strictement admissible mais pas d'un $Z^{(0)}$ strictement admissible

Tout d'abord, nous modifions quelque peu le problème primal par l'introduction d'une assez grande constante M que nous prenons comme borne supérieure de la trace de $F(x)$. Le problème (1.1) devient le problème modifié suivant

$$\min c^T x \quad \text{s.c. } F(x) \succeq 0, \quad \text{Retr} F(x) \leq M. \quad (5.21)$$

Si M est suffisamment grand, les solutions optimales de ce nouveau problème primal sont égales à celles du problème primal de départ. Le point $x^{(0)}$ strictement admissible pour (1.1) le restera si $\text{Retr} F(x^{(0)}) < M$.

Intéressons-nous ensuite au problème dual du problème modifié (5.21). Nous obtenons

$$\max -\text{Retr} F_0(Z - zI) - Mz \text{ s.c. } \text{Retr} F_i(Z - zI) = c_i \text{ pour } i = 1 \dots m, Z \succeq 0, z \geq 0. \quad (5.22)$$

Nous démontrons ci-dessous que $z^{(0)} > -\min(\lambda_{\min}(U), 0)$ où U est une matrice hermitienne telle que $\text{Retr} F_i U = c_i$ pour $i = 1 \dots m$ et $Z^{(0)} = U + z^{(0)}I$ forment un couple strictement admissible pour (5.22).

Nous avons $\text{Retr} F_i(Z^{(0)} - z^{(0)}I) = \text{Retr} F_i U = c_i$ pour $i = 1 \dots m$. De plus, $z^{(0)} > 0$, car, par définition, $z^{(0)} > -\min\{\lambda_{\min}(U), 0\} \geq 0$. Ensuite, nous avons que $Z^{(0)} \succeq 0$, c'est-à-dire que pour tout $z \in \mathbb{R}^n$, $z^T(U + z^{(0)}I)z = z^T U z + z^{(0)}\|z\|_2^2 \geq 0$, car, puisque U est une matrice hermitienne, $z^T U z \geq \lambda_{\min}(U)\|z\|_2^2 \geq \min(\lambda_{\min}(U), 0)\|z\|_2^2 > -z^{(0)}\|z\|_2^2$.

Il y a néanmoins deux inconvénients à cette méthode. Tout d'abord le choix de M n'est pas toujours aisé. Ensuite, nous sommes tenus de vérifier que la solution optimale du problème primal modifié trouvée ne dépende pas de M , c'est-à-dire que la contrainte $\text{tr} F(x) \leq M$ ne soit pas active en la solution optimale. Si c'est le cas, nous devons augmenter M et résoudre le nouveau problème modifié.

Connaissance d'un $Z^{(0)}$ strictement admissible mais pas d'un $x^{(0)}$ strictement admissible

Tout d'abord, nous modifions quelque peu le problème primal par l'introduction d'une assez grande constante M . Le problème (1.1) devient

$$\min c^T x + Mt \text{ s.c. } F(x) + tI \succeq 0, t \geq 0. \quad (5.23)$$

Intéressons-nous ensuite au problème dual du problème primal modifié (5.23). Nous obtenons

$$\max -\text{Retr} F_0 Z \text{ s.c. } \text{Retr} F_i Z = c_i \text{ pour } i = 1 \dots m, \text{Retr}(Z + z) = M, Z \succeq 0, z \geq 0. \quad (5.24)$$

qui est équivalent au problème dual de (1.1) si nous éliminons la variable d'écart z et si nous rajoutons la contrainte $\text{Retr} Z \leq M$. Nous obtenons

$$\max -\text{Retr} F_0 Z \text{ s.c. } \text{Retr} F_i Z = c_i \text{ pour } i = 1 \dots m, \text{Retr} Z \leq M, Z \succeq 0.$$

Nous démontrons ci-dessous que $x^{(0)}$ et $t^{(0)} > -\min(\lambda_{\min}(F(x^{(0)})), 0)$ forment un couple strictement admissible pour le problème primal modifié (5.23).

Nous avons $F(x^{(0)}) + t^{(0)}I \succeq 0$, c'est-à-dire que pour tout $z \in \mathbb{R}^n$, $z^T F(x^{(0)})z + t^{(0)}\|z\|_2^2 \geq 0$, puisque $F(x^{(0)})$ étant une matrice hermitienne par hypothèse, $z^T F(x^{(0)})z \geq \lambda_{\min}(F(x^{(0)}))\|z\|_2^2 \geq \min(\lambda_{\min}(F(x^{(0)})), 0)\|z\|_2^2 > -t^{(0)}\|z\|_2^2$.

De plus, nous avons $t^{(0)} > 0$ puisque $t^{(0)} > -\min(\lambda_{\min}(F(x^{(0)})), 0) \geq 0$.

Les inconvénients restent bien évidemment les mêmes. Le choix de M n'est point toujours évident et la solution optimale duale trouvée ne peut dépendre de la constante introduite. Si c'est le cas, nous sommes tenus d'augmenter M et de résoudre le problème modifié correspondant.

Connaissance ni d'un $Z^{(0)}$ strictement admissible ni d'un $x^{(0)}$ strictement admissible

Dans ce cas, nous combinons les deux méthodes développées dans les deux cas précédents en introduisant deux très grandes constantes, M_1 et M_2 . Le problème primal modifié est donc

$$\min c^T x + M_1 t \quad \text{s.c. } F(x) + tI \geq 0, \quad \text{Retr} F(x) \leq M_2, \quad t \geq 0,$$

et le dual modifié

$$\max -\text{Retr} F_0(Z - zI) - M_2 z \quad \text{s.c. } \text{Retr} F_i(Z - zI) = c_i \quad i = 1 \dots m, \quad \text{Retr} Z \leq M_1, \quad Z \succeq 0, \quad z \geq 0.$$

Nous savons, par les deux raisonnements précédents, que $x^{(0)}$ tel que $\text{Retr} F(x^{(0)}) \leq M_2$ et $t^{(0)} > -\min(\lambda_{\min}(F(x^{(0)})), 0)$ forment un point strictement admissible pour le primal tandis que $z^{(0)} > -\min(\lambda_{\min}(U), 0)$ où U est une matrice hermitienne telle que $\text{Retr} F_i U = c_i$ pour $i = 1 \dots m$ et $Z^{(0)} = U + z^{(0)}I$ tel que $\text{Retr} Z^{(0)} \leq M_1$ forment un point strictement admissible pour le dual.

Les inconvénients des deux cas précédents sont d'application dans ce cas-ci, ce qui rend cette méthode plus faible que celle basée sur la self-dualité. Cette dernière a le défaut d'être un peu plus compliquée et de supposer le cône \mathcal{K} solide, hypothèse peu restrictive.

5.3 Détection du caractère admissible ou non admissible du problème de départ

Les algorithmes présentés dans les chapitres 3 et 4, pour la programmation semi-définie, présupposent tous la forte admissibilité primale et duale. Cependant, à priori, nous ne pouvons juger du caractère admissible de ces deux problèmes initiaux. Nous tâchons dans cette section d'obtenir des informations sur l'admissibilité d'un problème primal de programmation conique convexe fermée et de son dual. Les résultats resteront valables pour la programmation semi-définie. Ainsi, en nous référant au tableau 2.2, nous aurons des précisions sur les valeurs optimales primale et duale. En nous restreignant à la programmation semi-définie, dans le cas où l'admissibilité forte primale-duale est respectée, nous pourrions appliquer la méthode primale-duale de points intérieurs et de suivi de chemin analysée dans les chapitres précédents.

Cette recherche se base sur l'existence de suites faiblement centrées, que nous définissons dans un premier temps. Celles-ci nécessitent la connaissance du problème de programmation conique convexe self-dual étendu (E) associé au problème primal-dual (SD), définie en (5.20). Les caractéristiques de (E) développées restent d'application.

Ensuite, nous démontrons le théorème qui, à partir d'une suite faiblement centrée de (E) associé à (SD), parvient à tirer des informations sur le caractère admissible du problème primal et du problème dual.

Dans un troisième temps, nous prouvons l'existence d'une suite faiblement centrée.

Mentionnons que l'hypothèse de la fermeture du cône est nécessaire pour obtenir un problème primal-dual qui soit self-dual.

Cette recherche est semblable à celle effectuée par Ye, Todd et Mizuno pour la programmation linéaire. A partir du problème primal-dual self-dual de départ, il y a création d'un problème extérieur qui permet de connaître l'admissibilité des problèmes primal et dual initiaux.

5.3.1 Définition d'une suite faiblement centrée

Nous définissons tout d'abord une suite faiblement centrée du problème (E) associé à (SD).

Une suite $(x_E^{(k)})_{k \in N} = ((x_{SD}^{(k)}, x_0^{(k)}, z_0^{(k)}))_{k \in N} \in (b_E + \mathcal{A}_E) \cap \mathcal{K}_H$ est dite faiblement centrée si et seulement si il existe une certaine constante $\omega \in]0, 1[$ telle que $x_0^{(k)} z_0^{(k)} \geq \omega c_E^T x_E^{(k)} > 0$ pour tout $k = 1, 2, \dots$ et $\lim_{k \rightarrow \infty} c_E^T x_E^{(k)} = 0$.

Dégageons immédiatement deux résultats caractérisant les suites faiblement centrées et nécessaires à la démonstration du théorème nous aidant à découvrir le caractère admissible ou non-admissible du problème primal et du problème dual de départ.

Lemme 5.3 Soit $(x_E^{(k)})_{k \in N}$ une suite faiblement centrée. Alors nous avons que

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{y_0^{(k)}}{x_0^{(k)}} = 0 & \iff \lim_{k \rightarrow \infty} z_0^{(k)} = 0 \\ \lim_{k \rightarrow \infty} \frac{y_0^{(k)}}{z_0^{(k)}} = 0 & \iff \lim_{k \rightarrow \infty} x_0^{(k)} = 0 \end{aligned}$$

preuve : Prouvons d'abord la première équivalence. Par définition d'une suite faiblement centrée, nous avons que

$$x_0^{(k)} z_0^{(k)} \geq \omega c_E^T x_E^{(k)} = \omega \frac{i^T \Pi_H i}{2} y_0^{(k)} > 0$$

par (5.14). Comme $x_0^{(k)}$ est strictement positif, nous avons

$$z_0^{(k)} \geq \omega \frac{i^T \Pi_H i}{2} \frac{y_0^{(k)}}{x_0^{(k)}} > 0.$$

Par conséquent, $\lim_{k \rightarrow \infty} z_0^{(k)} = 0$ implique $\lim_{k \rightarrow \infty} \frac{y_0^{(k)}}{x_0^{(k)}} = 0$.

De plus, si $\lim_{k \rightarrow \infty} \frac{y_0^{(k)}}{x_0^{(k)}} = 0$, nous avons $\lim_{k \rightarrow \infty} z_0^{(k)} = 0$, par la page 93 de [14].

La seconde équivalence résulte d'un raisonnement similaire.

cqfd

Le lemme suivant démontre une propriété importante des suites faiblement centrées. Les composantes $x_0^{(k)}$ et $z_0^{(k)}$ évitent le plus possible la frontière du cône R^+ .

Lemme 5.4 *Soit $x_E = (x, x_0, z_0) \in (b_E + \mathcal{A}_E) \cap \mathcal{K}_H$ et soit $\omega \in]0, 1[$ tel que $x_0 z_0 \geq \omega c_E^T x_E > 0$. Alors, pour tout $x'_E = (x', x'_0, z'_0) \in (b_E + \mathcal{A}_E) \cap \mathcal{K}_H$, nous avons*

$$x_0 \geq \frac{\omega}{1 + \frac{c_E^T x'_E}{c_E^T x_E}} x'_0 \text{ et } z_0 \geq \frac{\omega}{1 + \frac{c_E^T x'_E}{c_E^T x_E}} z'_0.$$

preuve : Prouvons tout d'abord la première inégalité. Puisque $x'_E - x_E \in \mathcal{A}_E$ et $\Pi_H(x'_E - x_E) \in \mathcal{A}_E^\perp$, nous obtenons successivement

$$\begin{aligned} 0 &= (x'_E - x_E)^T \Pi_H(x'_E - x_E) \\ &= x_E'^T \Pi_H x'_E + x_E^T \Pi_H x_E - x_E'^T \Pi_H x_E - x_E^T \Pi_H x'_E. \end{aligned}$$

Or, nous savons que

$$\begin{aligned} x_E'^T \Pi_H x_E &= x_E'^T \Pi_H^T x_E \text{ car } \Pi_H \text{ est symétrique} \\ &= (\Pi_H x'_E)^T x_E \\ &= x_E^T \Pi_H x'_E. \end{aligned}$$

Dès lors, nous avons successivement

$$\begin{aligned} 0 &= x_E'^T \Pi_H x'_E + x_E^T \Pi_H x_E - 2x_E'^T \Pi_H x'_E \\ &= 2c_E^T(x'_E + x_E) - 2(x^T \Pi_{SD} x' + x_0 z'_0 + z_0 x'_0) \end{aligned}$$

par le théorème 5.1. Tenant compte du fait que $x \in \mathcal{K}_{SD}$ et $\Pi_{SD} x' \in \mathcal{K}_{SD}^*$, ainsi que du fait que $x_0 z'_0 \geq 0$, nous avons

$$c_E^T(x'_E + x_E) \geq z_0 x'_0.$$

Nous multiplions les deux membres par $\frac{x_0}{c_E^T(x'_E + x_E)}$, ceci étant rendu possible puisque $c_E^T x_E > 0$ par hypothèse et $c_E^T x'_E \geq 0$ par le théorème 5.1. Enfin, nous utilisons la propriété $x_0 z_0 \geq \omega c_E^T x_E$ pour conclure que

$$x_0 \geq \frac{\omega}{1 + \frac{c_E^T x'_E}{c_E^T x_E}} x'_0.$$

La seconde inégalité résulte d'un raisonnement similaire.

cqfd

5.3.2 Recherche d'informations sur le caractère admissible du problème

Nous trouvons ci-dessous le théorème qui permet d'obtenir des informations sur le caractère admissible des problèmes primal et dual de départ en passant par l'analyse d'une suite faiblement centrée du problème de programmation conique convexe self-dual étendu (E) associé au problème primal-dual self-dual (SD).

Théorème 5.4 *Considérons un problème de programmation conique convexe fermée $CP(b, c, \mathcal{A}, \mathcal{K})$ et son dual. Ces deux problèmes nous permettent de définir le problème (SD) associé (voir (5.20)).*

Etant donné le problème (E) de programmation conique convexe self-dual étendu associé à (SD), soit $(x_E^{(k)})_{k \in \mathbb{N}}$ une suite faiblement centrée de (E).

Alors,

1. $\liminf_{k \rightarrow \infty} x_0^{(k)} > 0$ si et seulement si (SD) a une solution self-complémentaire,
2. $\liminf_{k \rightarrow \infty} z_0^{(k)} > 0$ si et seulement si (SD) est fortement inadmissible,
3. Si $\lim_{k \rightarrow \infty} z_0^{(k)} = 0$, alors
 - (a) $\lim_{k \rightarrow \infty} \text{dist}(\frac{x^{(k)}}{x_0^{(k)}}, b + \mathcal{A}) = 0$
 - (b) $\lim_{k \rightarrow \infty} \frac{c^T x^{(k)} + z_0^{(k)}}{x_0^{(k)}} \leq p^*$
 - (c) $d^* \geq -\liminf_{k \rightarrow \infty} \frac{c^T x^{(k)}}{x_0^{(k)}}$
 - (d) $d^* \geq -\limsup_{k \rightarrow \infty} \frac{c^T x^{(k)} + z_0^{(k)}}{x_0^{(k)}}$
4. Si $\lim_{k \rightarrow \infty} \frac{z_0^{(k)}}{x_0^{(k)}} = \infty$, alors (SD) est inadmissible et $\frac{x_{SD}^{(k)}}{z_0^{(k)}}$ est une suite de directions améliorantes dans \mathcal{K}_{SD} , c'est-à-dire

$$\begin{cases} \lim_{k \rightarrow \infty} \text{dist}(\frac{x_{SD}^{(k)}}{z_0^{(k)}}, \mathcal{A}_{SD}) = 0, \\ \lim_{k \rightarrow \infty} \frac{c_{SD}^T x_{SD}^{(k)}}{z_0^{(k)}} = -1 < 0. \end{cases}$$

preuve :

1. Nous savons que (SD) a une solution self-complémentaire x_{SD}
 - $\iff (x_{SD}, 1, 0)$ appartient à $\mathcal{A}_H \cap \mathcal{K}_H$ par (5.5)
 - $\iff (y_{SD}, x_0, 0)$ où $y_{SD} = x_{SD}x_0$ et $x_0 > 0$ est une direction non-nulle de $CP(0, 0, \mathcal{A}_H, \mathcal{K}_H)$
 - \iff (E) a une solution optimale $(x_{SD}^*, x_0^*, 0)$ avec $x_0^* > 0$ par le théorème 5.3.
- Par conséquent, il suffit de prouver l'équivalence entre $\liminf_{k \rightarrow \infty} x_0^{(k)} > 0$ et "(E) a une solution optimale de la forme $(x_{SD}^*, x_0^*, 0)$ avec $x_0^* > 0$ " pour obtenir la thèse.

D'une part, supposons que le problème (E) possède une solution optimale de la forme $x_E^* = (x_{SD}^*, x_0^*, 0)$ avec $x_0^* > 0$. Par la définition de suite faiblement centrée et par le

lemme 5.4, nous obtenons

$$x_0^{(k)} \geq \omega x_0^* \frac{1}{1 + \frac{c_E^T x_E^*}{c_E^T x_E^{(k)}}} = \omega x_0^* > 0 \quad \forall k$$

puisque $c_E^T x_E^* = 0$ et $c_E^T x_E^{(k)} \neq 0$. Par conséquent, $\liminf_{k \rightarrow \infty} x_0^{(k)} \geq \omega x_0^* > 0$.

D'autre part, supposons que $\liminf_{k \rightarrow \infty} x_0^{(k)} > 0$. Soit x_E^* un point d'adhérence de la suite $(x_E^{(k)})_{k \in N}$. Ce point existe puisque la suite $(x_E^{(k)})_{k \in N}$ est une suite bornée. En effet, une suite faiblement centrée est une suite bornée, (voir [14] page 95). De plus, par définition d'une suite faiblement centrée, nous avons $x_E^{(k)} \in (b_E + \mathcal{A}_E) \cap \mathcal{K}_H \forall k$, qui reste vrai lors du passage à la limite puisque (E) est un problème self-dual et donc \mathcal{K}_H est un cône fermé. Ensuite, puisque $\lim_{k \rightarrow \infty} c_E^T x_E^{(k)} = c_E^T x_E^* = 0$, par la définition de $y_0^{(k)}$, par le fait que $\liminf_{k \rightarrow \infty} x_0^{(k)} > 0$ et que $(x_E^{(k)})_{k \in N}$ est une suite bornée ainsi que par le lemme 5.3, nous trouvons que $z_0^* = 0$. De plus, puisque $c_E^T x_E^* = 0$, le théorème 4.1 disant que la fonction objectif de (E) est toujours positive nous conduit à affirmer que $x_E^* = (x_{SD}^*, x_0^*, 0)$ est une solution optimale de (E) avec $x_0^* > 0$.

2. Nous avons de façon similaire que (SD) est fortement inadmissible
 - \iff (SD) a une direction améliorante, par le tableau 2.1
 - \iff il existe x_{SD} tel que $x_{SD} \in \mathcal{A}_{SD} \cap \mathcal{K}_{SD}$ et $c_{SD}^T x_{SD} < 0$
 - \iff il existe x_{SD} et $z_0 = -c_{SD}^T x_{SD}$ tels que $(x_{SD}, 0, z_0) \in \mathcal{A}_H \cap \mathcal{K}_H$ avec $z_0 > 0$ par (5.4)
 - \iff (E) a une solution optimale $(x_{SD}^*, 0, z_0^*)$ avec $z_0^* > 0$, par le théorème 5.3
 - \iff $\liminf_{k \rightarrow \infty} z_0^{(k)} > 0$ par un même raisonnement que dans 1.
3. Montrons dans un premier temps que $\lim_{k \rightarrow \infty} z_0^{(k)} = 0$ entraîne que $\frac{x_{SD}^{(k)}}{x_0^{(k)}}$ est une suite dans \mathcal{K}_{SD} pour laquelle

$$\lim_{k \rightarrow \infty} \text{dist}\left(\frac{x_{SD}^{(k)}}{x_0^{(k)}}, b_{SD} + \mathcal{A}_{SD}\right) = 0 \quad (5.25)$$

$$\lim_{k \rightarrow \infty} \frac{c_{SD}^T x_{SD}^{(k)} + z_0^{(k)}}{x_0^{(k)}} = 0 \leq p_{SD}^*. \quad (5.26)$$

En effet, par le lemme 5.3, nous avons $\lim_{k \rightarrow \infty} \frac{y_0^{(k)}}{x_0^{(k)}} = 0$. De plus, puisque $x_E^{(k)} = (x_{SD}^{(k)}, x_0^{(k)}, z_0^{(k)}) \in (b_E + \mathcal{A}_E) \cap \mathcal{K}_E$, posant $i = (u_{SD}, u_0, v_0)$ et par (5.4), nous obtenons en réécrivant en partie la contrainte $x_E^{(k)} - y_0^{(k)}i \in \mathcal{A}_H \cap \mathcal{K}_H$ de (E),

$$x_{SD}^{(k)} - y_0^{(k)} u_{SD} \in (x_0^{(k)} - y_0^{(k)} u_0) b_{SD} + \mathcal{A}_{SD} \quad (5.27)$$

$$z_0^{(k)} - y_0^{(k)} v_0 = -c_{SD}^T (x_{SD}^{(k)} - y_0^{(k)} u_{SD}). \quad (5.28)$$

Divisant (5.27) par $x_0^{(k)}$ et faisant tendre k vers l'infini, nous obtenons

$$\lim_{k \rightarrow \infty} \frac{x_{SD}^{(k)}}{x_0^{(k)}} \in b_{SD} + \mathcal{A}_{SD} \quad (5.29)$$

qui est (5.25). Divisant (5.28) également par $x_0^{(k)}$ et faisant tendre k vers l'infini, nous obtenons

$$\lim_{k \rightarrow \infty} \frac{c_{SD}^T x_{SD}^{(k)} + z_0^{(k)}}{x_0^{(k)}} = 0 \leq p_{SD}^* \quad (5.30)$$

qui est (5.26).

Montrons dans un second temps (a). Par, (5.29) et par le fait que $x_{SD}^{(k)} = (x^{(k)}, z^{(k)})$, nous avons

$$\begin{aligned} \lim_{k \rightarrow \infty} \text{dist}\left(\frac{x^{(k)}}{x_0^{(k)}}, b + \mathcal{A}\right) &= 0 \quad (\text{a}) \text{ est prouv   et} \\ \lim_{k \rightarrow \infty} \text{dist}\left(\frac{z^{(k)}}{x_0^{(k)}}, c + \mathcal{A}^\perp\right) &= 0. \end{aligned}$$

Prouvons ensuite (b), (c) et (d). Par le 2., $\lim_{k \rightarrow \infty} z_0^{(k)} = 0$ entraine que (SD) n'est pas fortement inadmissible. Donc, ni le probl  me primal (P) ni le probl  me dual (D) ne sont fortement inadmissibles. Il suit du th  or  me 2.5 que $\bar{p} = -d^*$ et $p^* = -\bar{d}$. De plus, par d  finition, nous avons

$$\bar{p}_{SD} := \liminf_{\epsilon \rightarrow 0} \{c_{SD}^T x_{SD} \mid x_{SD} \in \mathcal{K}_{SD} \text{ dist}(x_{SD}, b_{SD} + \mathcal{A}_{SD}) < \epsilon\}.$$

Suite    cette d  finition, nous avons $\bar{p} \leq \liminf_{k \rightarrow \infty} \frac{c^T x^{(k)}}{x_0^{(k)}}$ et $\bar{d} \leq \liminf_{k \rightarrow \infty} \frac{b^T z^{(k)}}{x_0^{(k)}}$.

Or, de (5.30), il suit que $\lim_{k \rightarrow \infty} \frac{c^T x^{(k)} + b^T z^{(k)} + z_0^{(k)}}{x_0^{(k)}} = 0$. Par cons  quent, nous avons $d^* = -\bar{p} \geq -\liminf_{k \rightarrow \infty} \frac{c^T x^{(k)}}{x_0^{(k)}}$ et (c) est prouv  .

De plus, puisque $\frac{z_0^{(k)}}{x_0^{(k)}}$ est toujours positif, nous obtenons

$$\begin{aligned} d^* = -\bar{p} &\geq -\liminf_{k \rightarrow \infty} \frac{c^T x^{(k)} + z_0^{(k)}}{x_0^{(k)}} \\ &\geq -\limsup_{k \rightarrow \infty} \frac{c^T x^{(k)} + z_0^{(k)}}{x_0^{(k)}} \end{aligned}$$

qui prouve (d).

(b) est v  rifi   car nous avons l'  quivalence suivante

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{c^T x^{(k)} + z_0^{(k)}}{x_0^{(k)}} &= -\lim_{k \rightarrow \infty} \frac{b^T z^{(k)}}{x_0^{(k)}} \\ \iff \lim_{k \rightarrow \infty} \frac{c^T x^{(k)} + z_0^{(k)}}{x_0^{(k)}} &= -\liminf_{k \rightarrow \infty} \frac{b^T z^{(k)}}{x_0^{(k)}} \leq -\bar{d} = p^*. \end{aligned}$$

4. Supposons que $\lim_{k \rightarrow \infty} \frac{z_0^{(k)}}{x_0^{(k)}} = \infty$. Puisque $(x_E^{(k)})_{k \in N}$ est une suite born  e, $z_0^{(k)}$ ne peut tendre vers l'infini. Par cons  quent, $\lim_{k \rightarrow \infty} x_0^{(k)} = 0$ et donc $\lim_{k \rightarrow \infty} \frac{y_0^{(k)}}{z_0^{(k)}} = 0$

par le lemme 5.3. Divisant (5.27) par $z_0^{(k)}$ et faisant tendre k vers l'infini, nous avons $\lim_{k \rightarrow \infty} \text{dist}(\frac{x_{SD}^{(k)}}{z_0^{(k)}}, \mathcal{A}_{SD}) = 0$. De même divisant (5.28) par $z_0^{(k)}$, nous obtenons $\lim_{k \rightarrow \infty} \frac{c_{SD}^T x_{SD}^{(k)}}{z_0^{(k)}} = -1 < 0$. Par définition, $(\frac{x_{SD}^{(k)}}{z_0^{(k)}})_{k \in \mathbb{N}}$ est une suite de directions améliorantes, ce qui nous permet de conclure à l'inadmissibilité de (SD).

cqfd

Ce théorème démontre l'utilité des suites faiblement centrées pour la recherche d'informations sur le caractère admissible du problème primal et du problème dual.

Tout d'abord, si nous pouvons générer une suite faiblement centrée pour (E) telle que $\liminf_{k \rightarrow \infty} x_0^{(k)} > 0$, alors le problème self-dual (SD) est fortement admissible et possède une solution self-complémentaire $x_{SD}^* = (x^*, z^*)$ qui est une solution optimale pour (SD). Le primal et le dual sont donc tous les deux fortement admissibles. Nous pouvons leur appliquer une méthode de points intérieurs pour trouver une solution primale-duale optimale.

Si par contre la suite faiblement centrée générée est telle que $\liminf_{k \rightarrow \infty} z_0^{(k)} > 0$, alors le problème (SD) est fortement inadmissible. Par la preuve, nous voyons que cette inadmissibilité est due à l'existence d'une direction améliorante pour (SD), soit $x_{SD} = (x, z)$. Dans ce cas, nous avons $c_{SD}^T x_{SD} = c^T x + b^T z < 0$. Par conséquent, soit $c^T x$ et $b^T z$ sont tous les deux négatifs, soit une des deux quantités est négative et l'autre positive. Si nous nous trouvons dans le premier cas - $c^T x$ et $b^T z$ sont tous les deux strictement négatifs - z démontre l'inadmissibilité forte primale et x l'inadmissibilité forte duale (tableau 2.1). Si nous nous situons dans le second cas - $c^T x < 0$ et $b^T z \geq 0$ par exemple - x démontre l'inadmissibilité forte duale et démontre que (P) est soit inadmissible, soit non borné, par le chapitre 2.

Dans les autres cas, (SD) doit être soit faiblement admissible soit faiblement inadmissible. Si d'une part, la suite faiblement centrée générée est telle que $\lim_{k \rightarrow \infty} z_0^{(k)} = 0$, alors nous pouvons trouver une suite de points pour laquelle le nombre de contraintes de (SD) violées converge vers zéro et les valeurs objectifs correspondantes ne sont pas éloignées de la valeur optimale (p^*, d^*) . Si d'autre part, la suite faiblement centrée générée est telle que $\lim_{k \rightarrow \infty} \frac{z_0^{(k)}}{x_0^{(k)}} = \infty$, alors (SD) est inadmissible.

Les deux derniers cas n'apportent que très peu d'informations.

5.3.3 Existence de suites faiblement centrées

Le théorème précédant qui permet de détecter l'admissibilité ou non des problèmes primal et dual initiaux se base sur l'existence d'une suite faiblement centrée du problème de programmation conique convexe self-dual étendu (E). Pour que ce théorème soit d'application, il faut prouver l'existence d'une telle suite.

Nous élaborons les définitions nécessaires à la démonstration du théorème montrant l'existence d'une telle suite et ensuite effectuons la preuve de celui-ci sous certaines hypothèses du cône \mathcal{K} . Celles-ci seront satisfaites pour le cône semi-défini positif. Ainsi, ce théorème

est d'application pour la programmation semi-définie.

Tout d'abord définissons une barrière homogène ν -logarithmique pour un cône convexe \mathcal{K} fermé, solide et pointé.

$F : \text{int}\mathcal{K} \rightarrow R$ est une barrière homogène ν -logarithmique pour \mathcal{K} , cône convexe fermé, solide et pointé si F est une fonction strictement convexe deux fois différentiable sur $\text{int}\mathcal{K}$ telle que $F(x^{(i)})$ tend vers l'infini pour toute suite $(x^{(i)})_{i \in \mathbb{N}}$ dans $\text{int}\mathcal{K}$ qui converge vers la frontière de \mathcal{K} , et vérifie la propriété suivante, $F(tx) = F(x) - \nu \log t$, pour tout $x \in \text{int}\mathcal{K}$, $t > 0$ où $\nu \geq 0$ est un paramètre fixé.

Pour le cône $\mathcal{K} = \mathcal{H}_n^+$, la barrière la plus fréquemment utilisée est la barrière homogène n logarithmique, $F(X) = -\log \det(X)$, (voir proposition 5.4.5 dans [3]).

Définissons également la conjuguée d'une barrière $F(x)$ homogène ν -logarithmique pour \mathcal{K}

$$F^*(z) := \sup_{x \in \text{int}\mathcal{K}} \{(-z)^T x - F(x)\}. \quad (5.31)$$

$F^*(z)$ est alors une barrière homogène ν -logarithmique pour \mathcal{K}^* , voir théorème 2.4.4 dans [3]. De plus, nous avons $F^{**} = F$, par [12].

Terminons par définir une barrière homogène ν -logarithmique Π self-conjuguée. Soit \mathcal{K} un cône solide convexe tel que $\mathcal{K}^* = \Pi\mathcal{K}$ pour une certaine matrice Π de permutation symétrique. Une barrière homogène ν -logarithmique $F : \text{int}\mathcal{K} \rightarrow R$ pour \mathcal{K} est Π self-conjuguée si et seulement si $F^*(\Pi x) = F(x)$ pour tout $x \in \text{int}\mathcal{K}$.

Le théorème suivant prouve l'existence d'une suite faiblement centrée pour (E). En effet, il démontre l'existence de $x_E(\mu) = (x_{SD}(\mu), x_0(\mu), z_0(\mu))$, pour tout μ strictement positif, qui appartiennent à $(b_E + \mathcal{A}_E) \cap \mathcal{K}_E$ et tels que $x_0(\mu)z_0(\mu) = \mu = \frac{1}{\nu+1}c_E^T x_E(\mu)$ et $\lim_{\mu \rightarrow 0} c_E^T x_E(\mu) = 0$. Ainsi, $(x_E(\mu))_{\mu > 0}$ est une suite faiblement centrée pour (E).

Théorème 5.5 Soit $CP(b, c, \mathcal{A}, \mathcal{K})$, un problème de programmation conique convexe où \mathcal{K} est un cône convexe fermé, solide et pointé. Soient le problème self-dual associé (SD), voir (5.20), et le problème self-dual étendu associé (E), voir (5.6).

Soit $F : \text{int}\mathcal{K} \rightarrow R$ une barrière homogène ν -logarithmique pour \mathcal{K} .

Définissons

$$\begin{aligned} F_{SD}(x_{SD}) &:= F_{SD}((x, z)) = F(x) + F^*(z) \text{ où } (x, z) \in \mathcal{K} \times \mathcal{K}^* = \mathcal{K}_{SD}, \\ F_E((x_{SD}, x_0, z_0)) &:= F_{SD}(x_{SD}) - \log x_0 - \log z_0. \end{aligned}$$

Définissons également $\Phi_\mu(x_E) := (-c_E)^T x_E - \mu F_E(x_E)$ et posons

$$x_E(\mu) := \operatorname{argmax}\{\Phi_\mu(x_E) | x_E \in (b_E + \mathcal{A}_E) \cap \text{int}\mathcal{K}_E\}.$$

Alors

1. $F_{SD}((x, z))$ est une barrière homogène 2ν -logarithmique et Π_{SD} self-conjuguée pour \mathcal{K}_{SD} ,
2. $F_E((x_{SD}, x_0, z_0))$ est une barrière homogène $(2\nu + 2)$ logarithmique Π_H self-conjuguée pour \mathcal{K}_H ,
3. $x_E(\mu)$ est bien défini pour tout μ strictement positif,
4. Posant $z_E(\mu) := -\mu \nabla F_E(x_E(\mu))$, nous obtenons
 - (a) $z_E(\mu) = \operatorname{argmax}\{-b_E^T z_E - \mu F_E^*(z_E) \mid z_E \in (c_E + \mathcal{A}_E^\perp) \cap \operatorname{int}\mathcal{K}_H^*\}$,
 - (b) $x_E(\mu)^T z_E(\mu) = 2(\nu + 1)\mu$,
 - (c) $x_E(\mu) = -\mu \nabla F_E^*(z_E(\mu))$.
5. $x_0(\mu)z_0(\mu) = \mu = \frac{1}{\nu+1}c_E^T x_E(\mu)$ et $\lim_{\mu \rightarrow 0} c_E^T x_E(\mu) = 0$.

preuve :

1. Nous savons que $F(x)$ est une barrière homogène ν -logarithmique pour \mathcal{K} et que $F^*(z)$ en est une pour \mathcal{K}^* . Par conséquent, $F_{SD} : \operatorname{int}(\mathcal{K} \times \mathcal{K}^*) \rightarrow \mathbb{R}$ est une fonction strictement convexe, deux fois différentiable sur $\operatorname{int}(\mathcal{K} \times \mathcal{K}^*)$ et telle que $F_{SD}((x^{(i)}, z^{(i)}))$ tend vers l'infini pour toute suite $(x^{(i)}, z^{(i)})_{i \in \mathbb{N}}$ dans $\operatorname{int}(\mathcal{K} \times \mathcal{K}^*)$ qui converge vers la frontière de $\mathcal{K} \times \mathcal{K}^*$.

Il nous reste à montrer que $F_{SD}(t(x, z)) = F_{SD}(x, z) - 2\nu \log t$ pour tout $(x, z) \in \operatorname{int}(\mathcal{K} \times \mathcal{K}^*)$ et $t > 0$. En effet, nous avons

$$\begin{aligned} F_{SD}(t(x, z)) &= F(tx) + F^*(tz) \\ &= F(x) + F^*(z) - 2\nu \log t. \end{aligned}$$

Il est évident, par définition de Π_{SD} , que F_{SD} est Π_{SD} self-duale.

2. Puisque par le 1., $F_{SD}(x_{SD})$ est une barrière homogène ν -logarithmique et puisque $-\log x_0 - \log z_0$ est une fonction strictement convexe qui tend vers l'infini lorsque x_0 ou z_0 se rapproche de zéro, il nous reste à montrer que $F_E(tx_E) = F_E(x_E) - (2\nu + 2)\log t$ pour tout $x_E \in \operatorname{int}\mathcal{K}_E$ et $t > 0$ et que $F_E^*(\Pi_H x_E) = F_E(x_E)$ pour tout $x_E \in \operatorname{int}\mathcal{K}_E$ pour obtenir la thèse.

D'une part, nous avons successivement

$$\begin{aligned} F_E(tx_E) &= F_E(t(x_{SD}, x_0, z_0)) \\ &= F_{SD}(tx_{SD}) - \log(tx_0) - \log(tz_0) \\ &= F_{SD}(x_{SD}) - 2\nu \log t - \log x_0 - \log t - \log z_0 - \log t \\ &= F_E(x_E) - (2\nu + 2)\log t. \end{aligned}$$

D'autre part, soit $x_E \in \operatorname{int}\mathcal{K}_E$, nous avons successivement

$$\begin{aligned} F_E^*(\Pi_H x_E) &= \sup_{y \in \operatorname{int}\mathcal{K}_E} \{(-\Pi_H x_E)^T y - F_E(y)\} \\ &= \sup_{y_{SD} \in \operatorname{int}\mathcal{K}_{SD}, y_0 \in \mathbb{R}^{++}, g_0 \in \mathbb{R}^{++}} \{-(\Pi_{SD} x_{SD})^T y_{SD} - z_0 y_0 - x_0 g_0 \\ &\quad - F_{SD}(y_{SD}) + \log(y_0) + \log(g_0)\} \end{aligned}$$

Dès lors, après quelques manipulations de l'expression, nous trouvons que

$$\begin{aligned} F_E^*(\Pi_H x_E) &= F_{SD}(x_{SD}) - \log z_0 - \log x_0 \\ &= F_E(x_E). \end{aligned}$$

3. Soit μ strictement positif. Par hypothèse, nous avons $x_E(\mu) = \operatorname{argmax}\{\Phi_\mu(x_E) | x_E \in (b_E + \mathcal{A}_E) \cap \operatorname{int}\mathcal{K}_E\}$. Nous savons d'une part que $\Phi_\mu(x_E)$ est une fonction strictement concave puisque $\Phi_\mu(x_E) = -c_E^T x_E - \mu F_E(x_E)$ où $F_E(x_E)$ est une fonction strictement convexe. D'autre part, $\Phi_\mu(x_E)$ est une fonction bornée sur $(b_E + \mathcal{A}_E) \cap \operatorname{int}\mathcal{K}_E$ puisque

$$\begin{aligned} \Phi_\mu(x_E) &= -c_E^T x_E - \mu F_E(x_E) \\ &= -z_E^T x_E + z_E^T b_E - \mu F_E(x_E) \\ &\leq b_E^T z_E + \mu \sup_{x_E \in \operatorname{int}\mathcal{K}_E} \left(-\frac{z_E^T}{\mu} x_E - F_E(x_E) \right) \\ &= b_E^T z_E + \mu F_E^*\left(\frac{z_E}{\mu}\right). \end{aligned}$$

Il suit que Φ_μ atteint son maximum sur $(b_E + \mathcal{A}_E) \cap \operatorname{int}\mathcal{K}_E$ et que ce maximum est unique. Ainsi $x_E(\mu)$ est bien défini pour tout μ strictement positif.

4. Par définition, $x_E(\mu) = \operatorname{argmax}\{\Phi_\mu(x_E) | x_E \in (b_E + \mathcal{A}_E) \cap \operatorname{int}\mathcal{K}_E\}$. Des conditions d'optimalité du premier ordre, nous savons que $x_E(\mu)$ satisfait

$$\nabla \Phi_\mu(x_E(\mu)) = -c_E - \mu \nabla F_E(x_E(\mu)) \in \mathcal{A}_E^\perp.$$

Posant $z_E(\mu) := -\mu \nabla F_E(x_E(\mu))$, nous obtenons

$$z_E(\mu) \in c_E + \mathcal{A}_E^\perp. \quad (5.32)$$

Montrons que $z_E(\mu) \in \operatorname{int}\mathcal{K}_E^*$. En effet, de manière générale, nous avons

$$F_E^*(z_E) = (-z_E)^T x_E - F_E(x_E) < \infty \text{ si } z_E = -\nabla F_E(x_E), \quad (5.33)$$

puisque par définition $F_E^*(z_E) := \sup_{x_E \in \operatorname{int}\mathcal{K}_E} \{(-z_E)^T x_E - F_E(x_E)\}$. Annulant la dérivée de $(-z_E)^T x_E - F_E(x_E)$, nous trouvons $z_E := -\nabla F_E(x_E)$. De plus, $F_E^*(z_E) < \infty$ entraîne que $z_E \in \operatorname{int}\mathcal{K}_E^*$ par définition de F_E^* . Par conséquent, $-\nabla F_E(x_E(\mu)) \in \operatorname{int}\mathcal{K}_E^*$ et puisque μ est strictement positif $z_E(\mu) := -\mu \nabla F_E(x_E(\mu)) \in \operatorname{int}\mathcal{K}_E^*$.

Sur base de ce développement, prouvons le (c), c'est-à-dire que $x_E(\mu) = -\mu \nabla F_E^*(z_E(\mu))$. En fait, puisque, pour F une barrière homogène ν -logarithmique quelconque, $F(tx) = F(x) - \nu \log t$, nous avons pour celle-ci $\nabla F(tx)t = \nabla F(x)$. Cette relation entraîne que $\nabla F(\frac{x}{t})\frac{1}{t} = \nabla F(x)$ qui est équivalent à $\nabla F(\frac{x}{t}) = t \nabla F(x)$. Dans notre cas, nous obtenons

$$\nabla F_E^*\left(\frac{z_E(\mu)}{\mu}\right) = \mu \nabla F_E^*(z_E(\mu)).$$

Or, nous avons $\nabla F_E^*(\frac{z_E(\mu)}{\mu}) = -x_E(\mu)$ par définition de $z_E(\mu)$ et par (5.33). Par conséquent, nous avons

$$x_E(\mu) = -\mu \nabla F_E^*(z_E(\mu)). \quad (5.34)$$

Prouvons ensuite le (a), c'est-à-dire que $z_E(\mu) = \operatorname{argmax}\{-b_E^T z_E - \mu F_E^*(z_E) \mid z_E \in (c_E + \mathcal{A}_E^\perp) \cap \operatorname{int}\mathcal{K}_H^*\}$.

En effet, nous avons d'une part, $z_E(\mu) \in (c_E + \mathcal{A}_E^\perp) \cap \operatorname{int}\mathcal{K}_H^*$, par (5.32) et (5.33). D'autre part, remplaçant $x_E(\mu)$ par $-\mu \nabla F_E^*(z_E(\mu))$ dans $x_E(\mu) \in b_E + \mathcal{A}_E$, nous obtenons les conditions d'optimalité $-b_E - \mu \nabla F_E^*(z_E(\mu)) \in \mathcal{A}_E$ pour le problème de maximisation suivant :

$$\max\{-b_E^T z_E - \mu F_E^*(z_E) \mid z_E \in (c_E + \mathcal{A}_E^\perp) \cap \operatorname{int}\mathcal{K}_H^*\}.$$

Finalement, montrons le (b), c'est-à-dire que $x_E(\mu)^T z_E(\mu) = 2(\nu + 1)\mu$. En effet, pour une barrière homogène ν -logarithmique, nous avons $F(tx) = F(x) - \nu \log t$. Dérivant cette fois-ci par rapport à t , nous obtenons $\nabla F(tx)^T x = -\nu \frac{1}{t}$. Evaluant cette expression en $t = 1$, nous avons $\nabla F(x)^T x = -\nu$. Tenant compte de cette propriété, nous obtenons

$$z_E(\mu)^T x_E(\mu) = -\mu \nabla F_E(x_E(\mu))^T x_E(\mu) = (2\nu + 2)\mu. \quad (5.35)$$

5. Nous avons tout d'abord que $x_E(\mu) = \Pi_H z_E(\mu)$. En effet, d'une part, nous avons

$$x_E(\mu) := \operatorname{argmax}\{\Phi_\mu(x_E) \mid x_E \in (b_E + \mathcal{A}_E) \cap \operatorname{int}\mathcal{K}_E\}$$

où $\Phi_\mu(x_E) = -c_E^T x_E - \mu F_E(x_E)$.

D'autre part, $z_E(\mu)$ est l'unique argument qui maximise

$$-b_E^T z_E - \mu F_E^*(z_E) = \Phi_\mu(\Pi_H z_E)$$

sous contraintes que $z_E \in (c_E + \mathcal{A}_E^\perp) \cap \operatorname{int}\mathcal{K}_E$. Ainsi, nous écrivons

$$\begin{aligned} x_E(\mu) = \Pi_H z_E(\mu) &= -\mu \Pi_H \nabla F_E(x_E(\mu)) \\ &= -\mu \begin{pmatrix} \Pi_{SD} & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \nabla F_{SD}(x_{SD}(\mu))^T \\ -\frac{1}{x_0(\mu)} \\ -\frac{1}{z_0(\mu)} \end{pmatrix} \end{aligned}$$

Par conséquent, nous obtenons $x_0(\mu) = \frac{\mu}{z_0(\mu)}$ et

$$\begin{aligned} x_0(\mu) z_0(\mu) = \mu &= \frac{1}{2\nu + 2} x_E(\mu)^T z_E(\mu) \text{ par (5.35)} \\ &= \frac{1}{\nu + 1} c_E^T x_E(\mu) \end{aligned}$$

par le théorème 5.1. Ceci implique que $\lim_{\mu \rightarrow 0} c_E^T x_E(\mu) = 0$.

cqfd

Conclusion

La programmation semi-définie possède deux avantages non négligeables. D'une part, elle recouvre une grande classe de problèmes standards tels la programmation linéaire, la programmation quadratique ou l'optimisation convexe,... et par conséquent s'applique dans un nombre important de domaines comme la théorie du contrôle, l'optimisation combinatoire non convexe ou même les statistiques. (voir [15]). D'autre part, la programmation semi-définie est une extension de la programmation linéaire. Ainsi, de nombreux résultats tels les relations de dualité ainsi que de nombreuses méthodes de résolution telles les méthodes de points intérieurs s'y généralisent. Dans ce mémoire, nous avons tenté au mieux de mettre en évidence les liens existants entre la programmation linéaire et la programmation semi-définie.

Nous avons étudié de manière approfondie la méthode primale-duale de points intérieurs et de suivi de chemin pour la résolution d'un problème de programmation semi-définie.

Tout d'abord, nous avons porté notre choix sur une méthode primale-duale vu les liens étroits explicités dans le chapitre deux entre un primal et son dual et vu l'efficacité du point de vue numérique d'une telle méthode. Puis, nous avons souligné l'importance d'une méthode primale-duale de points intérieurs pour une progression rapide vers une solution optimale. Enfin, afin de faciliter l'obtention de la direction de Newton pour l'algorithme général de résolution d'un problème primal-dual de programmation semi-définie, nous avons eu recours au V-espace. Cet aller-retour entre l'espace de départ admissible et l'espace transformé a imposé à nos itérés de suivre approximativement une trajectoire spécifique. Ainsi, nous avons opté pour une méthode primale-duale de points intérieurs et de suivi de chemin. Dès lors, nous avons élaboré un algorithme général de résolution d'un problème de programmation semi-définie basé sur cette méthode. Ainsi, trois algorithmes relatifs à cette méthode et bien connus de la programmation linéaire ont pu être étendus à la programmation semi-définie. Nous avons supposé toutefois l'admissibilité forte primale-duale.

Vu le caractère récent de la programmation semi-définie, de nombreuses recherches sont encore menées sur le choix d'une direction de descente pour une telle méthode. Nous avons donné un bref aperçu des différentes manières de calculer une telle direction ainsi que quelques exemples. Nous avons ainsi établi des liens entre des directions différentes, entre autres, l'appartenance à une même famille de la direction de Monteiro-Tsuchiya et de la

direction de Nesterov-Todd développée dans ce mémoire.

Ensuite, nous avons démontré que tout algorithme découlant de l'algorithme général de résolution d'un problème de programmation semi-définie, basé sur la méthode primale-duale de points intérieurs et de suivi de chemin, converge vers le centre analytique de l'ensemble des solutions primales-duales optimales, similairement à la programmation linéaire. Nous avons établi la convergence des points situés sur la trajectoire centrale. Celle-ci nous a permis de prouver la convergence superlinéaire d'un algorithme de type prédicteur-correcteur.

Nous avons également proposé deux procédés, généralisés de la programmation linéaire, permettant de résoudre un problème de programmation semi-définie dont on ne connaît pas de point de départ strictement admissible. L'un se base sur la propriété de self-dualité et l'autre sur l'introduction d'une très grande constante positive M dans le problème initial. Finalement, nous avons pu, en étendant à la programmation conique convexe, la méthode de Ye, Todd et Mizuno, récolter des informations permettant de juger de l'admissibilité forte primale et duale ainsi que de l'inadmissibilité forte primale et duale. Les critères relevés dans ce mémoire caractérisant les autres cas d'admissibilité n'étant que très peu satisfaisant, il serait intéressant de pousser plus loin les recherches. N'oublions pas que ces résultats découlent de la dualité en programmation conique convexe que nous avons étendue de la programmation linéaire. Nous avons remarqué que la dualité s'y avère plus faible.

Nous concluons que cette méthode est satisfaisante. Non seulement, elle généralise la méthode de suivi de chemin établie en programmation linéaire et s'applique dans de nombreux domaines, mais elle possède également une complexité polynomiale et s'avère performante d'après les utilisateurs.

Annexe

Projections sur \mathcal{H} et sur \mathcal{H}^\perp

Etant donné une matrice carrée X , la projection de X sur \mathcal{H} vaut

$$\mathcal{P}_{\mathcal{H}}(X) = \frac{X + X^H}{2}$$

et la projection de X sur \mathcal{H}^\perp vaut

$$\mathcal{P}_{\mathcal{H}^\perp}(X) = \frac{X - X^H}{2}.$$

Produit de Kronecker

Soient $X, Z \in C^{n \times n}$ et

$$\begin{aligned} f_{X,Z} : \mathcal{H} &\longrightarrow \mathcal{H} \\ Y &\longrightarrow f_{X,Z}(Y) = \mathcal{P}_{\mathcal{H}}(ZYX^H). \end{aligned}$$

Puisque $f_{X,Z}(Y)$ est linéaire en Y , il existe une matrice, notée $X \otimes_H Z$ telle que

$$(X \otimes_H Z) \text{vec}_H Y = \text{vec}_H(f_{X,Z}(Y)) \quad \forall Y \in \mathcal{H}.$$

Le lecteur peut consulter la page 7 de [14].

Trace

Soient $M_n = C^{n \times n}$ et $T : M_n \longrightarrow M_n$ une application linéaire. Alors les assertions suivantes sont équivalentes :

1. il existe $S \in M_n$ telle que $T(X) = SXS^{-1}$, $\forall X \in M_n$, ou $T(X) = SX^T S^{-1}$, $\forall X \in M_n$.
2. $\det T(X) = \det X$ et $\text{tr} T(X) = \text{tr} X$ $\forall X \in M_n$.

Le lecteur peut consulter le point 4.5.7 page 293 de [4].

Normes et distance

La norme de Frobenius d'une matrice Y est donnée par $\|Y\|_F = \sqrt{Y \bullet Y}$.
 Pour Y une matrice hermitienne, nous avons $\|Y\|_F = \sqrt{\sum_{j=1}^n \lambda_j(Y)^2}$ où $\lambda_1(Y), \lambda_2(Y), \dots, \lambda_n(Y)$ sont les valeurs propres de Y .

Rappelons quelques propriétés des normes matricielles $\|\cdot\|_2$ et $\|\cdot\|_F$:

$$\begin{aligned} \|X\|_2 &\leq \|X\|_F \\ \|XY\|_F &\leq \|X\|_2 \|Y\|_F \\ \|X\|_2 &= \|X^H\|_2 \text{ et } \|X\|_F = \|X^H\|_F \end{aligned}$$

pour deux matrices quelconques X, Y .

O, o, θ

Soient $\{u(t)|t > 0\}$ et $\{w(t)|t > 0\}$ des suites réelles avec $w(t) > 0$. Nous posons les notations suivantes :

1. $u(t) = O(w(t))$ si et seulement si $\frac{u(t)}{w(t)}$ est bornée de manière indépendante de t .
2. $u(t) = o(w(t))$ si et seulement si $\lim_{t \rightarrow 0} \frac{u(t)}{w(t)} = 0$.
3. $u(t) = \Theta(w(t))$ si et seulement si $\frac{u(t)}{w(t)}$ et $\frac{w(t)}{u(t)}$ sont bornées de manière indépendante de t .

Pour une matrice $U(t) \in \mathcal{H}^{++}$, nous avons $U(t) = \Theta(w(t))$ signifie qu'il existe une constante $\Gamma > 0$ telle que

$$\frac{1}{\Gamma} I \preceq \frac{U(t)}{w(t)} \preceq \Gamma I$$

pour tout t strictement positif. Remarquons que si $U(t) = \Theta(w(t))$, alors nous avons $\|U(t)\| = \Theta(w(t))$ ainsi que $\lambda_i(U(t)) = \Theta(w(t))$ où $\lambda_i(U(t))$ désigne la $i^{\text{ème}}$ valeur propre de $U(t)$.

Bibliographie

- [1] F.Alizadeh, J.A.Haeberly, and M.Overton. A new primal-dual interior point method for semidefinite programming. In J.G.Lewis, editor, Proceedings of the Fifth SIAM Conference on Applied Linear Algebra, pages 113-117. SIAM, Philadelphia, 1994.
- [2] F.Alizadeh, J.A.Haeberly, and M.Overton. Primal-dual interior-point methods for semidefinite programming : convergence rates, stability and numerical results. Technical Report 721, Computer Science Department, New York University, New York, 1996.
- [3] G.H.Golub and C.F.van Loan. Matrix computations. The Johns Hopkins University Press. 1989.
- [4] Horn and Johnson. Topics in matrix analysis.
- [5] A.W.Marshall and I.Olkin. Inequalities : Theory of majorization and its applications. Academic Press 1979.
- [6] R.D.C. Monteiro and T.Tsuchiya. Polynomial convergence of a new family of primal-dual algorithms for semidefinite programming. Technical report, School of Industrial and Systems Engineering, Georgia Tech, Atlanta, Georgia, U.S.A., 1996.
- [7] R.D.C. Monteiro and P.R.Zanjácomo. A note on the existence of the Alizadeh-Haeberly-overton direction for semidefinite programming. Technical report, School of Industrial and Systems Engineering, Georgia Tech, Atlanta, Georgia, U.S.A., 1996. Revised January, 1997.
- [8] Y.Nesterov and A.Nemirovsky. Interior point polynomial methods in convex programming. SIAM, Philadelphia. 1994.
- [9] Y.Nesterov and M.J.Todd. Primal-dual interior-points methods for self-scaled cones. Technical Report 1125, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York, 1995.
- [10] Y.Nesterov and M.J.Todd. Self-scaled barriers and interior-point methods for convex programming. Mathematics of Operations Research, 22(1) :1-42, 1997.
- [11] J.M.Ortega and W.C.Rheinboldt. Iterative solution of nonlinear equations in several variables. Computer Science and Applied Mathematics Werner Rheinboldt University of Maryland. Academic Press. 1970.
- [12] R.T.Rockafellar. Convex Analysis. Princeton University Press, Princeton, NJ, 1970.
- [13] M.Shida, S.Shindoh, and M.Kojima. Existence of search directions in interior point algorithms for the SDP and the monotone SDLCP. Technical Report B-310, Dept. of

Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, 1996.

- [14] J.F.Sturm. Primal-dual interior point approach to semidefinite programming. Doctoral thesis. University of Rotterdam. September 1997.
- [15] L.Vandenberghe and S.Boyd. Semidefinite programming. SIAM Review. 1996.
- [16] Y.Zhang. On extending primal-dual interior-point algorithms from linear programming to semidefinite programming. Technical report, Dept. of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, Maryland, U.S.A., 1995.

Table des matières

Introduction	4
1 La programmation semi-définie	6
1.1 Trois formulations du problème de programmation semi-définie	6
1.2 Applications de la programmation semi-définie	10
1.2.1 Programmation quadratique avec contraintes quadratiques	10
1.2.2 Minimisation de la valeur propre maximale d'une matrice et de la norme matricielle	10
1.2.3 Schéma de séparation par ellipsoïde	11
1.2.4 Optimisation combinatoire et non convexe	12
1.2.5 Théorie du contrôle	13
2 La dualité	15
2.1 Le dual	15
2.2 Terminologie	16
2.3 Propriétés de base des cônes convexes	18
2.4 Caractérisation de l'admissibilité et de l'inadmissibilité	20
2.5 Dualité faible et dualité forte	26
3 Méthodes de suivi de chemin en programmation semi-définie	31
3.1 Le V-espace	32
3.1.1 Transformation du problème initial	32
3.1.2 Choix de la matrice de transformation L pour un point intérieur primal-dual (X, Z)	35
3.1.3 Avantages de la matrice de transformation L_d et trajectoire centrale	38
3.1.4 Calcul de L_d et de V	40
3.1.5 Choix de la matrice de transformation $G(t)$ de points intérieurs primaux-duaux $(X(t), Z(t))$	40
3.2 Calcul de la direction de Newton en un point intérieur primal-dual	41
3.2.1 Calcul de la direction de Newton en un point intérieur primal-dual $(\bar{X}(0), \bar{Z}(0))$ de l'espace transformé	42
3.2.2 Calcul de la direction de Newton en un point primal-dual (X, Z) de l'espace initial	46
3.3 Voisinages de la trajectoire centrale	47

3.4	Lemmes techniques	49
3.5	Algorithmes de suivi de chemin	49
3.5.1	Algorithme général de suivi de chemin	50
3.5.2	Algorithme à petits pas	50
3.5.3	Algorithme prédicteur-correcteur	52
3.5.4	Algorithme au plus grand pas	55
3.5.5	Algorithme à longs pas	57
3.6	Directions de descente en programmation semi-définie pour les méthodes primales-duales de points intérieurs et de suivi de chemin	59
3.6.1	Recherche d'une direction de descente basée sur les barrières dites "self-scaled"	59
3.6.2	Recherche d'une direction de descente basée sur les cibles matricielles	62
3.6.3	Recherche d'une direction de descente basée sur les cibles pour les valeurs propres de XZ	64
4	Analyse de la convergence de la méthode primale-duale de points intérieurs et de suivi de chemin pour la programmation semi-définie	70
4.1	Centre analytique de l'ensemble des solutions primales-duales optimales . .	71
4.1.1	Partition d'une matrice hermitienne	71
4.1.2	μ -centre et centre analytique	72
4.2	Convergence des points de la trajectoire centrale primale-duale vers le centre analytique de l'ensemble des solutions optimales primales-duales	75
4.3	Analyse de la convergence d'un algorithme de type prédicteur-correcteur .	81
4.3.1	Algorithme $SDP(\epsilon)$ de type prédicteur-correcteur	81
4.3.2	Convergence globale de l'algorithme $SDP(\epsilon)$	84
4.3.3	Convergence superlinéaire du saut de dualité vers 0 ainsi que de $(X^{(k)}, Z^{(k)})$ vers le centre analytique (X^a, Z^a)	87
5	Recherche d'informations sur le caractère admissible du problème ainsi que d'un itéré de départ intérieur dans le cas fortement admissible	92
5.1	La self-dualité	93
5.2	Recherche d'un point intérieur initial $(x^{(0)}, z^{(0)})$	95
5.2.1	Méthode basée sur la self-dualité : cas d'un problème de program- mation conique convexe self-dual	95
5.2.2	Méthode basée sur la self-dualité : extension au cas de la program- mation conique convexe fermée	103
5.2.3	La méthode du grand M	104
5.3	Détection du caractère admissible ou non admissible du problème de départ	106
5.3.1	Définition d'une suite faiblement centrée	107
5.3.2	Recherche d'informations sur le caractère admissible du problème .	109
5.3.3	Existence de suites faiblement centrées	112
	Conclusion	117

Annexe	119
Projections sur \mathcal{H} et sur \mathcal{H}^\perp	119
Produit de Kronecker	119
Trace	119
Normes et distance	120
O, o, θ	120
Bibliographie	121